Analysis of various data mining classification techniques to predict diabetes mellitus

[1]T.Nithyapriya, [2] S.Dhinakaran

[1]M.Phil. Scholar, , [2]Assistant Professor
[1]Department of Computer Science,
[1]Rathinam College of Arts and Science, Coimbatore, India

Abstract—Data mining approach helps to diagnose patient's diseases. Diabetes Mellitus is a chronic disease to affect various organs of the human body. Early prediction can save human life and can take control over the diseases. This research work explores the early prediction of diabetes using various data mining techniques. The real time diabetic based dataset has taken with 203 instances for training data set and 52 instances for test data set to determine the accuracy of the Naïve Bayes, SVM and J48 classification techniques in prediction. The analysis proves that SVM Classifier provide the highest accuracy than other techniques.

Index Terms—Data mining, Diabetes, Prediction, accuracy, classification

I. INTRODUCTION

Data mining take holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. Some experts believe the opportunities to get better care and reduce costs simultaneously could apply to as much as 30% of overall healthcare spending. This could be a win/win overall. But due to the complexity of healthcare and a slower rate of technology adoption, our industry lags behind these others in implementing effective data mining and analytic strategies.

Data mining has become an essential methodology for computing applications in medical informatics. Progress in data mining applications and its implications are manifested in the areas of information management in healthcare organizations, health informatics, epidemiology, patient care and monitoring systems, assistive technology, large-scale image analysis to information extraction and automatic identification of unknown classes. Various algorithms associated with data mining have significantly helped to understand medical data more clearly, by distinguishing pathological data from normal data, for supporting decision-making as well as visualization and identification of unseen complex relationships between diagnostic features of different patient groups.

II. DATA MINING IN DIABETIC MELLITUS

Diabetes mellitus, or simply diabetes, is a set of related diseases in which the body cannot regulate the amount of sugar in the blood [4]. It is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced. This high blood sugar make the classical symptoms of polyuria, polydipsia and polyphagia [5]. There are three main types of diabetes mellitus (DM). Type 1 DM results from the body's failure to produce insulin, and presently requires the person to inject insulin or wear an insulin pump. This form was before referred to as "insulin dependent diabetes mellitus" (IDDM) or "juvenile diabetes". Type 2 DM results from insulin resistance, a condition in which cells fail to use insulin properly, sometimes combined with an absolute insulin deficiency. This form was previously referred to as noninsulin dependent diabetes mellitus (NIDDM) or "adult-onset diabetes". The third main form, gestational diabetes occurs when pregnant women without a previous diagnosis of diabetes develop a high blood glucose level. It may lead development of type 2 DM.

Data Mining [6] refers to extract or mining knowledge from huge amounts of data. The aim of data mining is to make sense of huge amounts of mostly unsupervised data, in some domain. Classification [1] maps data into predefined groups. It is often referred to as supervised learning as the classes are determined prior to examining the data. Classification Algorithms usually require that the classes be defined based on the data attribute values. They often describe these classes by looking at the characteristics of data already known to belong to class. Pattern Recognition is a type of classification where an input pattern is classified into one of the several classes based on its similarity to these predefined classes. Knowledge Discovery in Databases (KDD) is the process of finding useful information and patterns in data which involves Selection, Pre-processing, Transformation, Data Mining and Evaluation.

Diabetic mellitus in India

In 2000, India (31.7 million) became topped the world with the highest number of people with diabetes mellitus followed by China (20.8 million) and United States (17.7 million) with diabetes mellitus[3]. According to Wild et al[2] the prevalence of diabetes is predicted to twice globally from 171 million in 2000 to 366 million in 2030 with a maximum increase in India. In 20130, 79.4 million individuals in India will be affected by diabetic mellitus, while China (42.3 million) and the United States (30.3 million) will also increases in those affected by the disease.

Nowadays, Indian became the diabetes capital of the world with as many as 50 million people suffering from type-2 diabetes, India has a challenge to face. However, medical specialists feel that timely detection and right management can go a long way in helping patients lead a normal life.

India having the highest number of diabetic patients in the world, the sugar disease is posing an enormous health problem to our country today including Pressure, time taken to heal the wounds, tiredness, blurred vision, etc., Often known as the diabetes capital of the world, India has been observing an alarming rise in incidence of diabetes according to the International Journal of Diabetes in Developing Countries. According to a World Health Organization's diabetes fact sheet, an estimated 34 lakhs deaths are caused due to high blood sugar.

Purpose of Study:

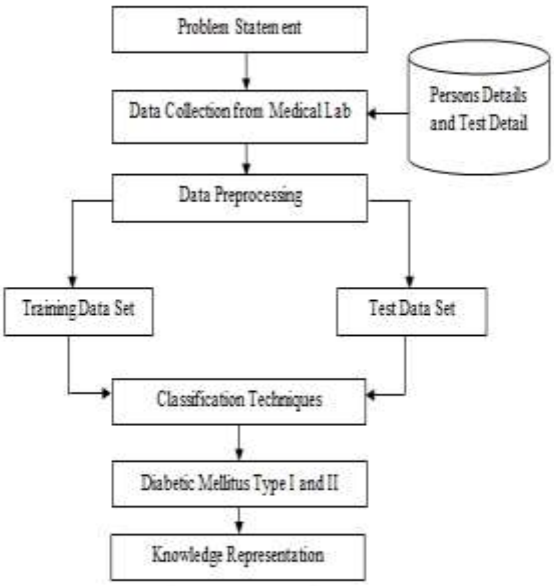
This study has aims to implement several prediction based classification techniques in data mining to assist medical institutions, medical research centres and labs with predicting the people's diabetic mellitus status. If the persons are predicted to have a chance to affected by the diabetic mellitus, then extra efforts can be made to improve their health conditions and allows to suggest the necessary steps to be taken to protect their health from diabetic mellitus.

III. MEDICAL DATA MINING USING CLASSIFICATION

Medical Data mining refers to extracting or "mining" knowledge from large amounts of medical data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. Currently, the data are stored in diabetic database, these database contain the useful information to predict diabetic mellitus. The most useful data mining techniques in medical database is classification.

We have Figure 3.1 that represents working methodology based on the framework. It is important to have a working methodology to govern our work before applying data mining techniques. The work methodology begins with problem definition, data collection and data preprocessing that includes data selection and data transformation and it precedes with data mining classification techniques with pruning which leads to discovering knowledge that is benefit to us.

Figure 3.1. Data mining work methodology



Data Set

Data collection questionnaire consists of 17 questions with sub-questions such as Name, Age, Weight, Physical activity , Urination, Water consumption, Diet, Systolic blood pressure, Hyper tension, Tiredness, Blurred vision, Wound healing, Sleepy/drowsy, Sudden weight loss, Heredity, Glucose level and Diabetic Mellitus presented.

Total size of the data set is 255 with 17 attributes. Collected all details are stored in Excel spreadsheet file (xls) format. It is used to predict the diabetic mellitus in the test data set using classification techniques.

Classification methods used

In this research work the following classification methods are used to predict the diabetic mellitus and also analyse the performance of these classification techniques in the diabetic data set

- Naïve Bayes
- Support Vector Machine
- J48

© 2017 IJEDR | Volume 5, Issue 4 | ISSN: 2321-9939

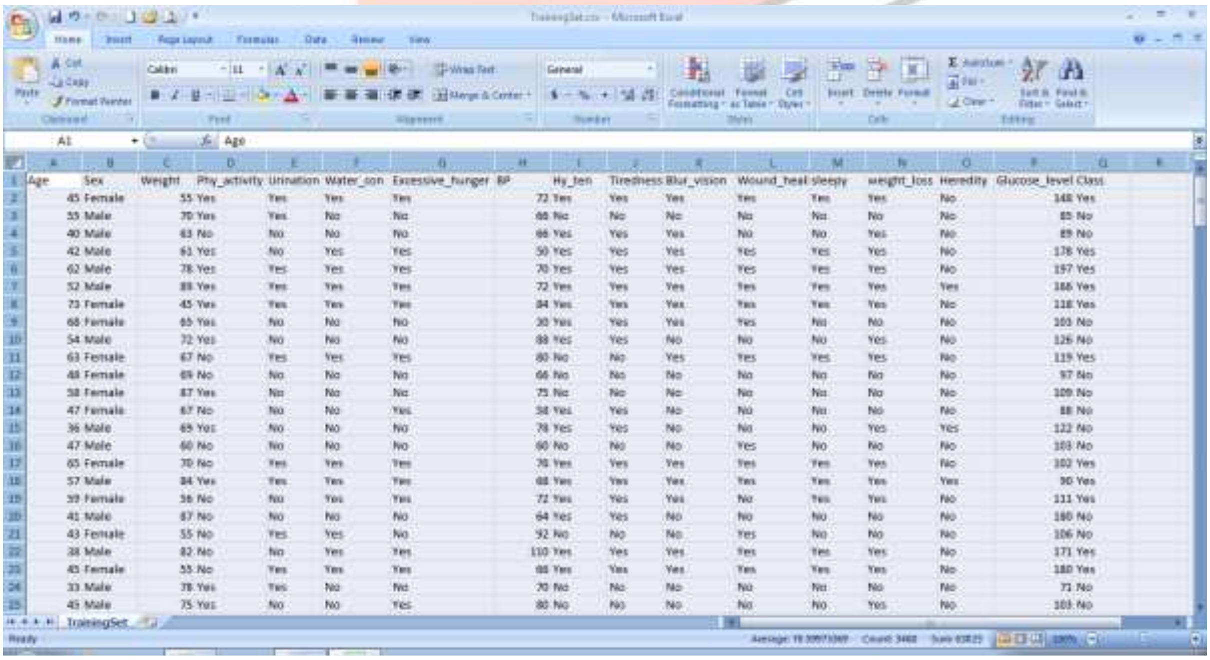
Attribute Selection

In those fields were chosen which were requisite for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table 3.1 for reference.

Table 3.1 Selected attributes

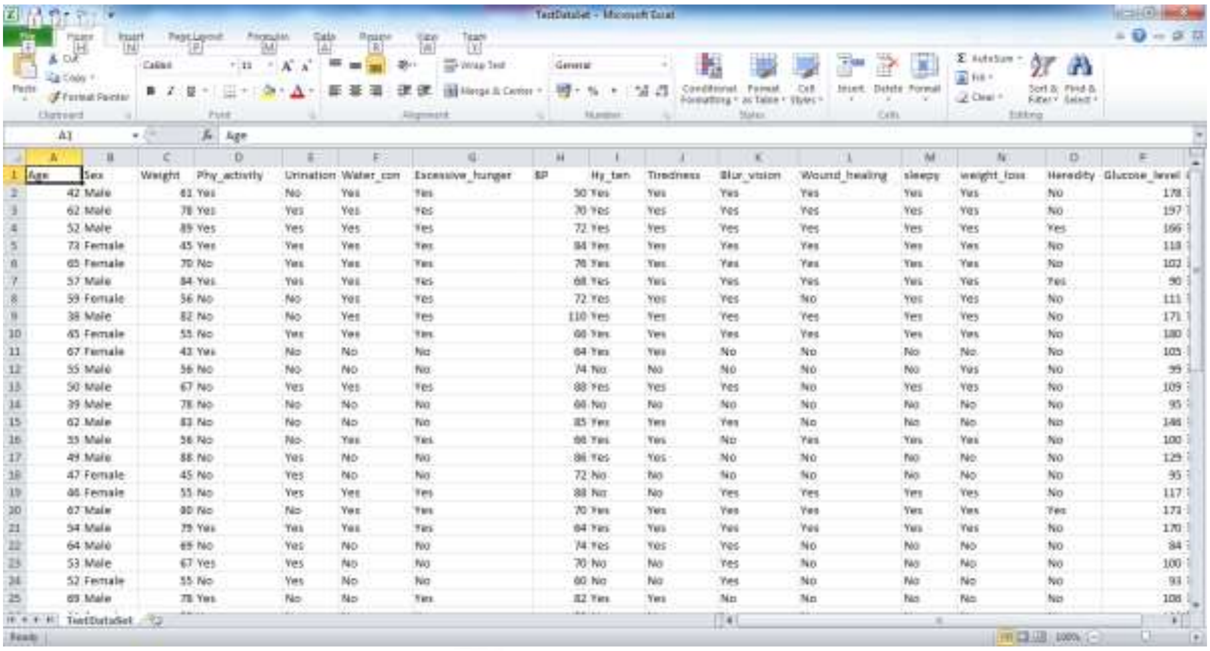
Variables	Description	Possible Values
Age	Age in years	{1 to 100}
Weight	Weight in Kg's	{5 to 120}
Physical activity	Physical activity in minutes	{Yes, No}
Urination	Number of times urination in a day	{Yes, No}
Water consumption	Water consumption in litres	{Yes, No}
Excessive Hunger	Excessive hunger in day time	{Yes, No}
Systolic blood pressure	Enter value of blood pressure in "mmHg"	{50 to 200}
Hyper tension	Person with hyper tension	{Yes, No}
Tiredness	Feel tiredness	{Yes, No}
Blurred vision	Have blurred vision	{Yes, No}
Wound healing	Wound Healing quickly	{Yes, No}
Sleepy/drowsy	Always feel sleepy/drowsy	{Yes, No}
Sudden weight loss	Observed sudden weight loss	{Yes ,No}
Heredity	Elders found with diabetes	{Yes, No}
Glucose level	Level of glucose in blood(No)	{50 to 400}
Diabetic	Diabetic Present	{Yes, No}

Fig.3.2. Training Data Set



Age	Sex	Weight	Phy_activity	Urination	Water_con	Excessive_hunger	BP	Hy_ten	Tiredness	Blur_vision	Wound_heal	sleepy	weight_loss	Heredity	Glucose_level	Class
45	Female	55	Yes	Yes	Yes	Yes	72	Yes	Yes	Yes	Yes	Yes	Yes	No	148	Yes
55	Male	70	Yes	Yes	No	No	66	No	No	No	No	No	No	No	85	No
40	Male	63	No	No	No	No	66	Yes	Yes	Yes	No	No	Yes	No	89	No
42	Male	61	Yes	No	Yes	Yes	50	Yes	Yes	Yes	Yes	Yes	Yes	No	178	Yes
62	Male	78	Yes	Yes	Yes	Yes	70	Yes	Yes	Yes	Yes	Yes	Yes	No	197	Yes
52	Male	88	Yes	Yes	Yes	Yes	72	Yes	Yes	Yes	Yes	Yes	Yes	Yes	188	Yes
73	Female	45	Yes	Yes	Yes	Yes	84	Yes	Yes	Yes	Yes	Yes	Yes	No	118	Yes
68	Female	65	Yes	No	No	No	30	Yes	Yes	Yes	Yes	No	No	No	103	No
54	Male	72	Yes	No	No	No	88	Yes	Yes	No	No	No	Yes	No	126	No
63	Female	67	No	Yes	Yes	Yes	80	No	No	Yes	Yes	Yes	Yes	No	119	Yes
48	Female	69	No	No	No	No	66	No	No	No	No	No	No	No	97	No
58	Female	87	Yes	No	No	No	75	No	No	No	No	No	No	No	109	No
47	Female	67	No	No	No	Yes	58	Yes	Yes	No	No	No	No	No	88	No
36	Male	69	Yes	No	No	No	78	Yes	Yes	No	No	No	Yes	Yes	122	No
47	Male	60	No	No	No	No	60	No	No	No	Yes	No	No	No	103	No
65	Female	70	No	Yes	Yes	Yes	78	Yes	Yes	Yes	Yes	Yes	Yes	No	102	Yes
57	Male	84	Yes	Yes	Yes	Yes	68	Yes	Yes	Yes	Yes	Yes	Yes	Yes	90	Yes
59	Female	56	No	No	Yes	Yes	72	Yes	Yes	Yes	No	Yes	Yes	No	111	Yes
41	Male	67	No	No	No	No	64	Yes	Yes	No	No	No	No	No	180	No
43	Female	55	No	Yes	Yes	No	92	No	No	No	Yes	No	No	No	106	No
38	Male	82	No	No	Yes	Yes	110	Yes	Yes	Yes	Yes	Yes	Yes	No	171	Yes
45	Female	55	No	Yes	Yes	Yes	68	Yes	Yes	Yes	Yes	Yes	Yes	No	180	Yes
33	Male	78	Yes	Yes	No	No	70	No	No	No	No	No	No	No	71	No
45	Male	75	Yes	No	No	Yes	80	No	No	No	No	No	Yes	No	103	No

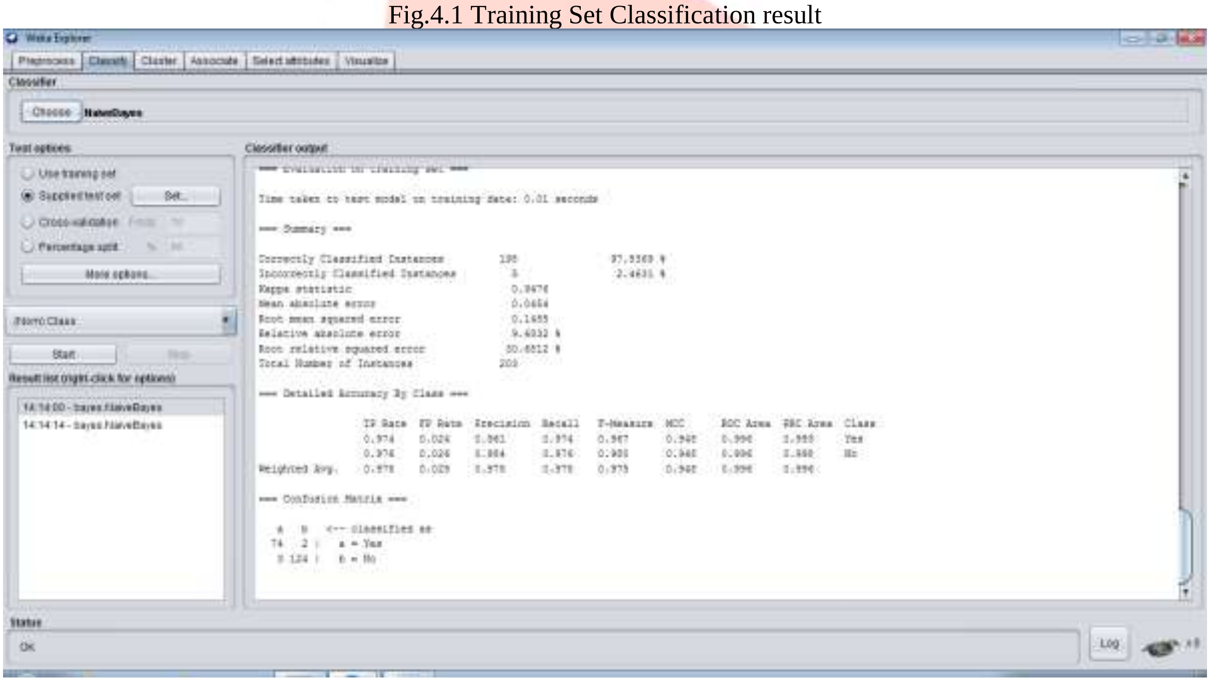
Fig.3.3. Test Data Set



IV. CLASSIFICATION RESULTS

4.3.1. NAÏVE BAYES CLASSIFICATION

Fig.4.1 Training Set Classification result



Naive Bayes doesn't select any important features. The result of the training of a Naive Bayes classifier is the mean and variance for every feature. The classification of new samples into 'Yes' or 'No' is based on whether the values of features of the sample match best to the mean and variance of the trained features for either 'Yes' or 'No' for the diabetic class variable.

In this test data 97.5369% of diabetic training data instances are correctly classified and remaining 2.4631% of diabetic instances are incorrectly classified. The percentage of correctly classified instances is often called accuracy or sample accuracy. So this data set consists of 97.5% accurate instances. The raw numbers are shown in the confusion matrix, with a and b representing the class labels. Here there were 203 instances, so the percentages and raw numbers add up, aa + bb = 74+124 = 198, ab + ba = 3+2 = 5. Kappa is a chance-corrected measure of agreement between the classifications and the true classes.

TP Rate: rate of true positives (instances correctly classified as a given class). Weighted average TP Rate of this data set is 0.975.

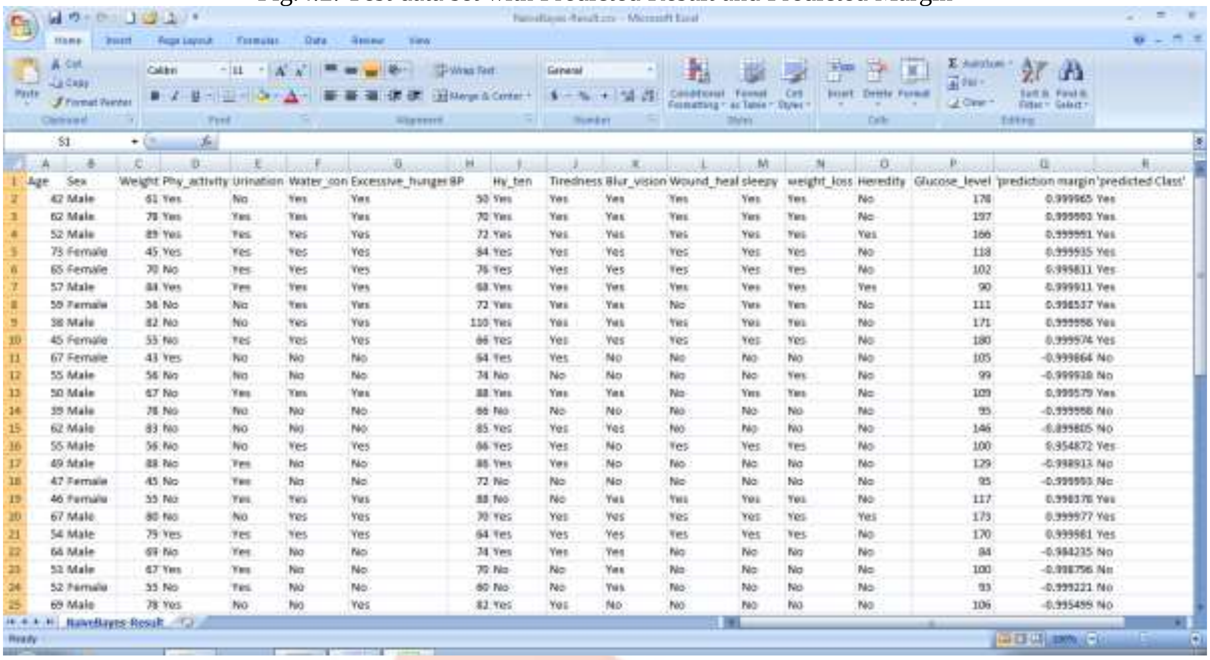
FP Rate: rate of false positives (instances falsely classified as a given class). Weighted average FP Rate of this data set is 0.025.

Precision: proportion of instances that are truly of a class divided by the total instances classified as that class. Weighted average Precision value of this data set is 0.975.

Recall: proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate). Weighted average Recall of this data set is 0.975.

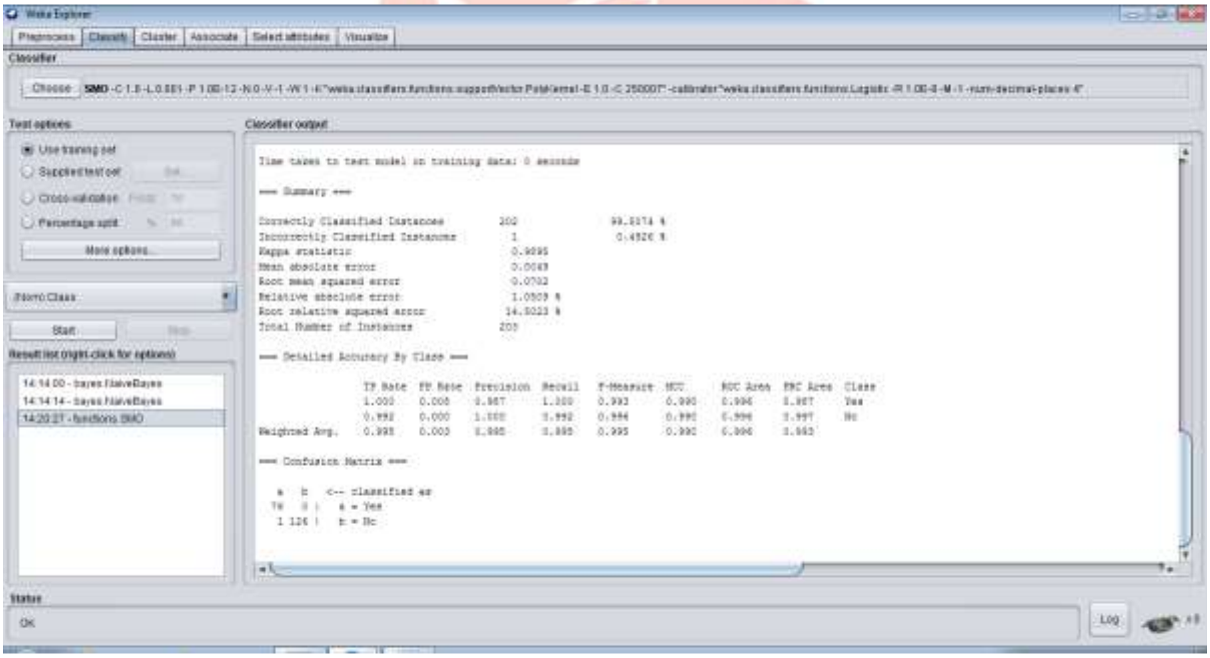
F-Measure: A combined measure for precision and recall calculated as 2 * Precision * Recall / (Precision + Recall). Weighted F-Measure value is 0.975.

Fig.4.2. Test data set with Predicted Result and Predicted Margin



4.3.2. SVM CLASSIFICATION

Fig.4.4. SVM Classifier result for diabetic data set



The correctly and incorrectly classified instances show the percentage of training instances that were correctly and incorrectly classified. In this test data 99.5074% of training data instances are correctly classified and remaining 0.4926% of instances are incorrectly classified. The percentage of correctly classified instances is often called accuracy or sample accuracy. So this data set consists of 99.5% accurate instances. The raw numbers are shown in the confusion matrix, with a and b representing the class labels. Here there were 203 instances, so the percentages and raw numbers add up, aa + bb = 76+126 = 202, ab + ba = 1+0 = 1. Kappa is a chance-corrected measure of agreement between the classifications and the true classes. It's calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. A value greater than 0 means that this classifier is doing better than chance.

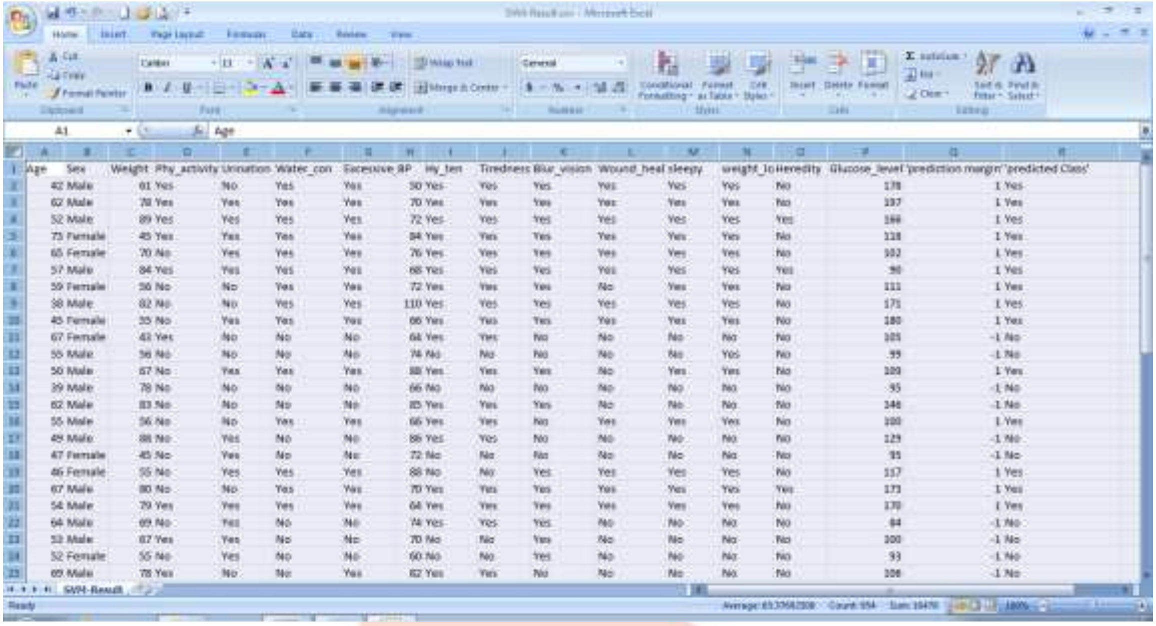
TP Rate: rate of true positives (instances correctly classified as a given class). Weighted average TP Rate of this data set is 0.995.

FP Rate: rate of false positives (instances falsely classified as a given class). Weighted average FP Rate of this data set is 0.003.

Precision: proportion of instances that are truly of a class divided by the total instances classified as that class. Weighted average Precision value of this data set is 0.995.

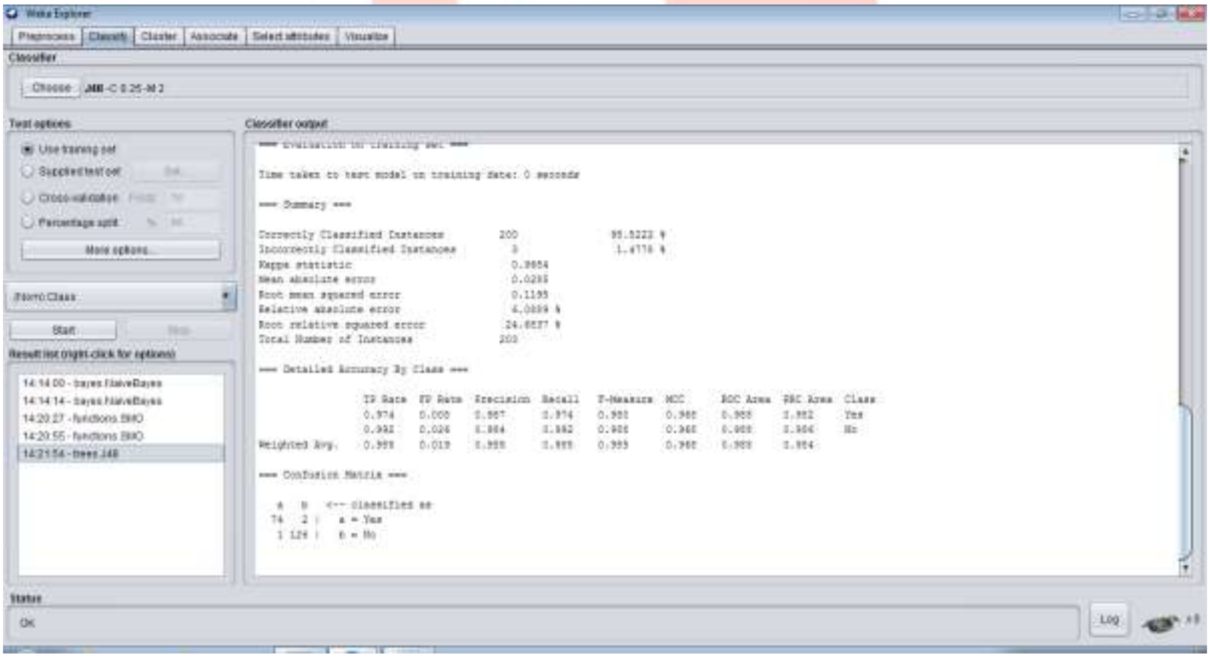
Recall: proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate). Weighted average Recall of this data set is 0.995.
F-Measure: A combined measure for precision and recall calculated as 2 * Precision * Recall / (Precision + Recall). Weighted F-Measure value is 0.995.

Fig.4.5. Test data set with Predicted Result and Predicted Margin by SVM



4.3.2. J48 CLASSIFICATION

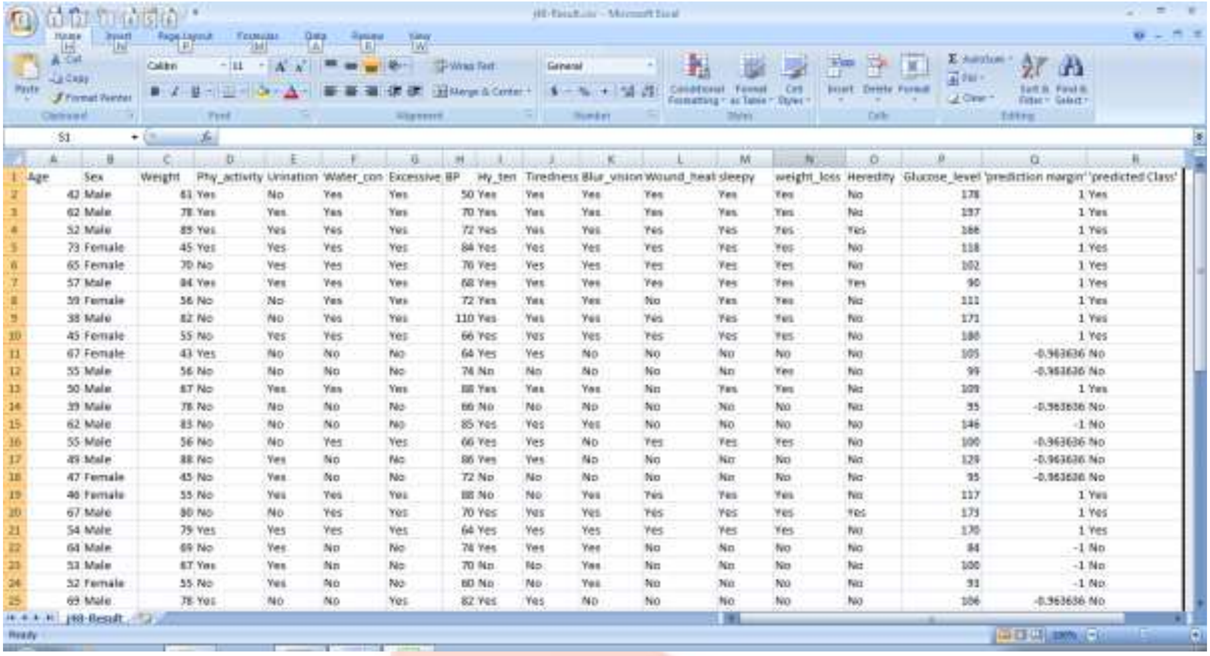
Fig.4.5. J48 Classification result for diabetic data set



In this diabetic test data 98.5222% of training data instances are correctly classified and remaining 1.4778% of instances are incorrectly classified. The percentage of correctly classified instances is often called accuracy or sample accuracy. So this data set consists of 98.5% accurate instances. The raw numbers are shown in the confusion matrix, with a and b representing the class labels. Here there were 203 instances, so the percentages and raw numbers add up, aa + bb = 74+126 = 200, ab + ba = 1+2 = 3.
TP Rate: rate of true positives (instances correctly classified as a given class). Weighted average TP Rate of this data set is 0.985.
FP Rate: rate of false positives (instances falsely classified as a given class). Weighted average FP Rate of this data set is 0.019.
Precision: proportion of instances that are truly of a class divided by the total instances classified as that class. Weighted average Precision value of this data set is 0.985.
Recall: proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate). Weighted average Recall of this data set is 0.985.

F-Measure: A combined measure for precision and recall calculated as 2 * Precision * Recall / (Precision + Recall). Weighted F-Measure value is 0.985.

Fig.4.6. Test data set with Predicted Result and Predicted Margin by J48



	Age	Sex	Weight	Phy_activity	Urination	Water_con	Excessive	BP	Hy_ten	Tiredness	Blur_vision	Wound_heal	sleepy	weight_loss	Heredity	Glucose_level	'prediction margin'	'predicted Class'
2	42	Male	61	Yes	No	Yes	Yes	50	Yes	Yes	Yes	Yes	Yes	Yes	No	178	1	Yes
3	62	Male	78	Yes	Yes	Yes	Yes	70	Yes	Yes	Yes	Yes	Yes	Yes	No	197	1	Yes
4	52	Male	89	Yes	Yes	Yes	Yes	72	Yes	Yes	Yes	Yes	Yes	Yes	Yes	166	1	Yes
5	73	Female	45	Yes	Yes	Yes	Yes	84	Yes	Yes	Yes	Yes	Yes	Yes	No	118	1	Yes
6	65	Female	70	No	Yes	Yes	Yes	76	Yes	Yes	Yes	Yes	Yes	Yes	No	102	1	Yes
7	57	Male	84	Yes	Yes	Yes	Yes	68	Yes	Yes	Yes	Yes	Yes	Yes	Yes	90	1	Yes
8	59	Female	56	No	No	Yes	Yes	72	Yes	Yes	Yes	No	Yes	Yes	No	111	1	Yes
9	38	Male	82	No	No	Yes	Yes	110	Yes	Yes	Yes	Yes	Yes	Yes	No	171	1	Yes
10	45	Female	55	No	Yes	Yes	Yes	66	Yes	Yes	Yes	Yes	Yes	Yes	No	180	1	Yes
11	67	Female	43	Yes	No	No	No	64	Yes	Yes	No	No	No	No	No	105	-0.963636	No
12	55	Male	56	No	No	No	No	74	No	No	No	No	No	Yes	No	99	-0.963636	No
13	50	Male	87	No	Yes	Yes	Yes	88	Yes	Yes	Yes	No	Yes	Yes	No	109	1	Yes
14	39	Male	78	No	No	No	No	66	No	No	No	No	No	No	No	35	-0.963636	No
15	62	Male	83	No	No	No	No	85	Yes	Yes	Yes	No	No	No	No	146	-1	No
16	55	Male	56	No	No	Yes	Yes	66	Yes	Yes	No	Yes	Yes	Yes	No	100	-0.963636	No
17	49	Male	88	No	Yes	No	No	86	Yes	Yes	No	No	No	No	No	129	-0.963636	No
18	47	Female	45	No	Yes	No	No	72	No	No	No	No	No	No	No	35	-0.963636	No
19	46	Female	55	No	Yes	Yes	Yes	88	No	No	Yes	Yes	Yes	Yes	No	117	1	Yes
20	67	Male	80	No	No	Yes	Yes	70	Yes	Yes	Yes	Yes	Yes	Yes	Yes	173	1	Yes
21	54	Male	79	Yes	Yes	Yes	Yes	64	Yes	Yes	Yes	Yes	Yes	Yes	No	170	1	Yes
22	64	Male	69	No	Yes	No	No	78	Yes	Yes	Yes	No	No	No	No	84	-1	No
23	53	Male	67	Yes	Yes	No	No	70	No	No	Yes	No	No	No	No	100	-1	No
24	52	Female	55	No	Yes	No	No	60	No	No	Yes	No	No	No	No	93	-1	No
25	69	Male	78	Yes	No	No	Yes	82	Yes	Yes	No	No	No	No	No	106	-0.963636	No

V. RESULTS AND DISCUSSIONS

5.1. COMPARISON OF CLASSIFICATION ALGORITHMS BASED ON CLASSIFIED INSTANCE

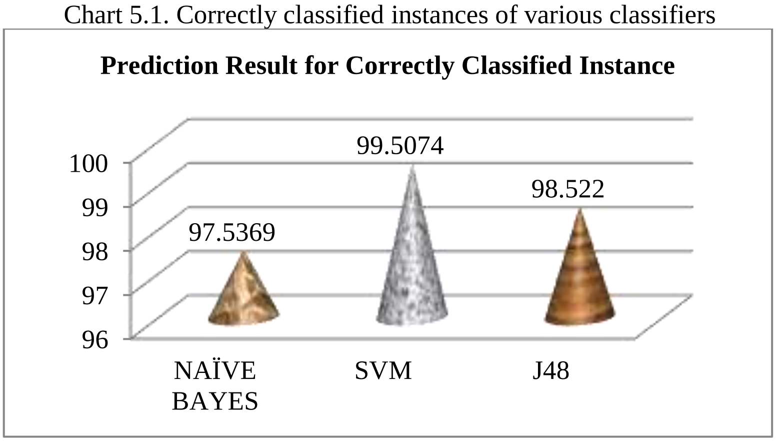
Table 5.1. Correctly classified instances of various classifiers

CLASSIFICATION ALGORITHM	CORRECTLY CLASSIFIED INSTANCE
NAÏVE BAYES	198
SVM	202
J48	200

The above table reveals that the out of 203 instances, 202 instances are correctly classified by the Support vector machine, 200 instances are correctly classified by the J48 classifier and 198 instances are correctly classified by the Naïve Bayes.

Support Vector Machine produced highest accuracy (99.50%) in the classification of diabetic data set. J48 classifier produced 98.52% accuracy and Naïve Bayes Classifier Produced 97.54% accuracy in the classification of diabetic data set.

Chart 5.1. Correctly classified instances of various classifiers



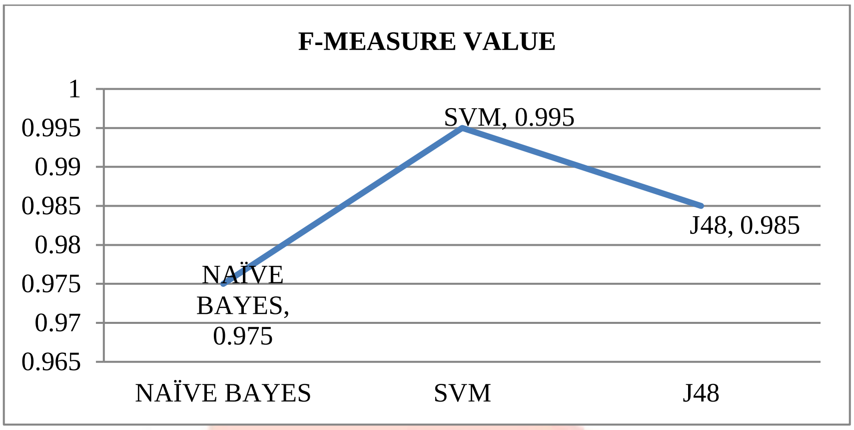
5.2. F-MEASURE VALUES FOR EACH CLASSIFICATION MODEL

Table 5.2. F-Measure values of various classifiers

CLASSIFICATION MODEL	F-MEASURE VALUE
NAÏVE BAYES	0.975
SVM	0.995
J48	0.985

The above table reveals that the classification process in the 203 instances, F-Measure value of Support Vector Machine is 0.995, J48 classifier's F-Measure value is 0.985 and Naïve Bayes classifier's F-Measure value is 0.975.

Chart 5.2. F-Measure values of various classifiers



5.3. COMPARISON MODELS

To calculate the performance of the various classification models, Correct Classified Rate (CCR), Recall Rate (RR), and F-measure to be used in document or data set classification criteria. The CCR is the rate of correct prediction, and Recall Rate is the ratio actually hit accurate predictions. And F-measure means the combinational mean of CCR and RR, and this is convenient expression method to compare models. The accuracy (AC) is the proportion of the total number of predictions that were correct. The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified. The false positive rate (FP) is the proportion of negatives cases that were incorrectly classified as positive.

Support Vector Machine classification model has the best F-measure value (0.995), Correctly classified rate (99.5%), best Recall rate (0.995) and best correctly classified instances (202) model showed the value compared to the other models such as Naïve Bayes and J48 decision tree in data classification of the diabetic data set with 203 instances. J48 classification model is closely followed to the Support vector machine model.

VI. CONCLUSION AND FUTURE ENHANCEMENT

Data Classification is an important application area in prediction mining in the medical data sets why because classifying millions of patient's records manually is an expensive and time consuming task. Therefore, automatic classifier is constructed using pre classified sample diabetic data set whose accuracy and time efficiency is much better than manual classification and prediction. Identifying efficient patterns also plays major role in text classification. Data mining classification techniques need to be designed to effectively manage large numbers of elements with varying frequencies. Almost all the known techniques for classification such as decision trees rules, Bayes methods and SVM classifiers have been used to the case of diabetic data.

In this research work, training and test diabetic data sets are used to predict the diabetic mellitus using various classification techniques. And we compared those data by applying the material to the conventional techniques of Bayesian statistical classification, J48 Decision tree and SVM to form a prediction model. The SVM model shows better performance than J48 and Naïve Bayes classification models. Future works may also include hybrid classification models by combining some of the data mining techniques such as attribute selection and clustering.

REFERENCES

[1] Han, J., Kamber, M.: Data Mining; Concepts and Techniques, Morgan Kaufmann Publishers (2000).
[2] Kumar A, Goel MK, Jain RB, Khanna P, Chaudhary V. India towards diabetes control: Key issues. Australas Med J. 2013;6(10):524–31.
[3] http://timesofindia.indiatimes.com/life-style/health-fitness/health-news/India-is-the-diabetes-capital-of-the-world/articleshow/50753461.cms
[4] http://www.emedicinehealth.com/diabetes.
[5] http://en.wikipedia.org/wiki/Diabetes_mellitus.
[6] http://diabetes.co.in.

[7] Margaret H. Dunham,-“Data Mining Techniques and Algorithms”, Prentice Hall Publishers.
[8] Danielle M.Hessler, lawrence Fisher, Joseph T.Mullan, Russel E.Glasgow, Umesh Masharani., Patient age: A neglected factor when considering disease management in adults with type 2 diabetes, *Elsevier Journal*,85 (2011), 154–159.
[9] Mira Kania Sabariah, Aini Hanifa and Siti Sa’adah, Early Detection of Type-2 Diabetes Mellitus with Random Forest, Classification and Regression Tree, *IEEE Transaction*, 2014, 238-242.
[10] L.Xu,C.Q.Jiang,C.M.Schooling,W.S.Zhang,K.K.Cheng,T.H.Lam., Prediction of 4-year incident diabetes in older chinese: Recalibration of Framingham diabetes score on Guangzhou Biobank Cohort Study, *Elsevier Journal*, 69 (2014), 63–68.
[11] Janice.M.S.Lopez, Robert.A.Bailey, Marcia.F.T.Rupnow, KathyAnnunziata., 2014, Characteri-zation of type2 diabetes Mellitus Burden by age and ethnic Groups Based on a Nationwide Survey, *Elsevier Journal*,36 (2014).
[12] M.Mounika et al, “Predictive Analysis of Diabetic Treatment Using Classification Algorithm” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2502-2505.
[13] Santi Wulan Purnami, S.P. Rahayu and AbdullahEmbong, “Feature selection and classification of breast cancer diagnosis based on support vector machine”, IEEE 2008.
[14] M. Durairaj, V. Ranjani, “Data Mining Applications In Healthcare Sector: A Study”, International journal of scientific & technology research volume 2, issue 10, October 2013, ISSN 2277-8616.
[15] P.Yasodha, M. Kannan, “Analysis of a Population of Diabetic Patients Databases in WEKA Tool”, International Journal of Scientific & Engineering Research Volume 2, Issue 5, May-2011, ISSN 2229-5518.
[16] Rashedur M. Rahman, Farhana Afroz, “Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis”, Journal of Software Engineering and Applications, 2013, 6, 85-97
[17] D.Lavanya, K.Usha Rani, “Performance Evaluation of Decision Tree Classifiers on Medical Datasets”, International Journal of Computer Applications, July 2011 (0975 – 8887) Volume 26– No.4.
[18] K. R. Lakshmi, S.Prem Kumar, “Utilization of Data Mining Techniques for Prediction of Diabetes Disease Survivability”, International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013, 933 ISSN 2229-5518.
[19] Akash Rajak, “A Temporal Reasoning System for Diagnosis and Therapy Planning”, I.J. Information Technology and Computer Science, 2015, 12, 23-29 Published Online November 2015 in MECS (http://www.mecs-press.org/) DOI:10.5815/ijitcs.2015.12.03
[20] Vaishali Jain , Supriya Raheja “Improving the Prediction Rate of Diabetes using Fuzzy Expert System” I.J. Information Technology and Computer Science, 2015, 10, 84-91 Published Online September 2015 in MECS
[21] Gao, Denzinger J, James RC. CoLe: A cooperative data mining approach and its application to early diabetes detection. Proceedings of the 5th International Conference on Data Mining (ICDM’05); 2005
[22] Rajesh K, Sangeetha V. Application of data mining methods and techniques for diabetes diagnosis. International Journal of Engineering and Innovative Technology (IJEIT). 2012; 2(3):224–9.
[23] AfrandP,Yazdani NM, Moetamedzadeh H, NaderiF,Panahi MS. Design and implementation of an expert clinical system for diabetes diagnosis. Global Journal of Science, Engineering and Technology; 2012. p. 23–31. ISSN:2322-2441.
[24] Adidela DR, Lavanya DG, Jaya SG, Allam AR. Application of fuzzy ID3 to predict diabetes. Int J AdvComput Math Sci. 2012; 3(4):541–5.
[25] PatilBM, Joshi RC, Toshniwal D. Association rule for classification of type-2 diabetic patients. 2nd International Conference of IEEE on Machine Learning and Computing; 2010. p. 67. DOI 10.1109/ICMLC.
[26] AljarullahAA. Decision tree discovery for the diagnosis of type II diabetes. International Conference on Innovative in Information Technology; 2011. p. 303–7.
[27] Jaya Rama Krishnaiah VV, Chandra Shekar DV, Satya Prasad R, Rao KRH. An empirical study about type-2 diabetes suing duo mining approach. International Journal of Computational Engineering Research. 2012; 2(6):33–42.
[28] Mandal S, Dubey V. Implementation and evaluation of diabetes management system using clustering technique. Special Issue of International Journal of Computer Science and Informatics. 2(2):33–6.
[29] Kavitha K, Sarojamma RM. Monitoring of diabetes with data mining via CART Method. International Journal of Emerging Technology and Advanced Engineering. 2012; 2(11):157–62.