DETERMINATION OF HEALTH CARE DISEASE USING DATA MINING TECHNIQUES

[1]Dr. Divya R. Jariwala, [2]Samiksha H. Zaveri, [3]Dr. Manish M. Kayasth
[1]Assistant Professor,[2]Research Scholar,[3]Assistant Professor and Head
[1,3]UCCC & SPBCBA & SDHG College of BCA and IT,Surat, Gujarat, India
[2]Parul University, Vadodara, Gujarat, India

Received: April 02, 2019 Accepted: May 11, 2019

ABSTRACT: *Data mining is the best technique for the enormous amounts of data processed and analyzed. Several data mining methodologies are existed and useful to transform these amount of data into useful information for making certain decision. Different classification data mining techniques produce different results for the same data set. Thus, finding the optimal algorithm for the given data set is a challenging task. The paper focuses on healthcare area where the aim is to analyzed diabetes patients via different supervised machine learning methods. The paper also shows an effective technique for detection of the diabetes disease where the outcome can be useful in selecting most suitable classifier for the given dataset. To determine the effectiveness of various classification algorithms like Linear Regression, Multilayer Perception, Gaussian Process, Simple Linear Regression, SMOreg, authors run some well-known classification algorithms against some standard datasets. Effectiveness of various algorithms is measured on various parameters like average accuracy, time taken to build classification model, mean absolute error, etc. Based on the comparative study of the experiment results, authors suggest the optimal algorithm.*

Key Words: *Linear Regression Model, Multilayer Perceptron, Gaussian Processes, Simple Linear Regression, SMOreg*

I. Introduction

Diabetes is a chronic condition that occurs when the human body cannot produce enough use of insulin[15]. The high blood sugar is due to two reasons. The first is either the pancreas does not produce enough insulin;or another reason is cells do not respond to the insulin that is produced [16].

When the glucose level in body is not metabolized properly, it keeps on circulating in the blood and causes damage to various tissues. High bold glucose level in diabetes is associated with long-term damage and dysfunction, and failure of various organs, especially the kidneys, eyes, nerves, heart, and blood vessels. Diabetes can mainly be of 3 types: type-1 diabetes, type-2 diabetes, and gestational diabetes. Type-1 diabetes results from nonproduction of insulin.Type-2 diabetes results from the development of resistance of insulin, as a result of which insulin produced is not able to metabolize the sugar levels properly. Gestational diabetes occurs in pregnant women, who develop a high blood glucose level during pregnancy and who never had any previous such history. It may be preceded by the development of type-2 diabetes [15].

Recognizing a type of diabetes to an individual is mostly based on the situation at the time of diagnosis. It is observed that most of the diabetic patients do not just fix into a single category [17] [18].

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information that can be used for industrial, medical and scientific purposes [16]. Data mining methods is one of the best ways to utilize large volumes of available diabetes-related data for extracting knowledge and making decisions [19].At present the results of early prediction of diabetes are not highly accurate. Therefore, there is a need to develop a diagnosis system for diabetes with better accuracy.

II. Data Mining In Diabetic Mellitus

Data mining involves computational process, machine learning, statistical techniques, classification, clustering and discovering patterns processes to extract useful information from large volume of database [20].

Medical data miningis used in the knowledge acquisition and analyses the information obtained from research reports, medical reports, flow charts, evidence tables, and transform these mounds of data into useful information for decision making.The present work has taken up to analyze the obtained data of

diabetic patients by various data mining classification algorithms which can be helpful for medical analysts or practitioners for accurate diabetes diagnosis [16].

Our current work mainly focuses on type-2 diabetic patients. The patient information of type-2 diabetes perspective used to find out classification algorithm's accuracy and error rate (ER) [21].This has done using Weka comprehensive classification techniques like linear regression model, Gaussian Process, Multilayer Perception, Simple Linear Regression and SMOreg.

This paper presents an analytical study on the existing techniques available for diabetes mellitus. The characteristic of the approaches are investigated to develop a better approach for the early and efficient diagnosis of the disease[18].

This study implement classification techniques in data mining to assist medical institutions, medical research centers and laboratory with predicting the human beings diabetic mellitus status. If the persons are predicted to have a chance to affect by the diabetic mellitus, then extra exertion can be made to improve their health conditions and suggesting the necessary steps to be taken to protect their health [22][23]. There are numerous factors causing diabetics in men and women are getting treated.

III. Literature Review

DeeptiSisodiaa, Dilip Singh Sisodia studied to design a model which can prognosticate the likelihood of diabetes in patients with maximum accuracy. Therefore three machine learning classification algorithms namely Decision Tree, SVM and Naive Bayes are used. The performances of all the three algorithms are evaluated on various measures like Precision, Accuracy, F-Measure, and Recall. Accuracy is measured over correctly and incorrectly classified instances. Results obtained show Naive Bayes performs with the highest accuracy of 76.30% compare with other algorithms [1].

Rafiqul Islam, ObaidurRahman, find out the risk factors of Type 2 diabetic patients in Bangladesh.To fulfill this objective, chi square test and logistic regression analysis have been used. It is found that diabetes affects more in the age 35 years and over which is 89.7% in which 68.3% have type 2 diabetes. Again, 79.3% of diabetic patients have type 2 diabetes in which females (43.7%) are more affected than males (35.7%). It has been found that age, controlling diabetic through exercise, controlling diabetic through taking medicine and living house of the respondents are significantly associated with the type 2 diabetes of diabetic patients [2].

S.UmmugulthumNatchiar, S.Baulkani, observedthat various computational intelligent techniques are used for diabetes classification. Among those rules based classification algorithms are highly used by the researchers for medical diagnosis. Since these decision rules are easily interpretable and understandable. They suggested that for decision rule generation roughset theory can also be used to handle noisy, missing and uncertainty data [3].

Koushik Chandra Howlader, Md. ShahriareSatu y, AvijitBarua, Mohammad Ali Moni, evaluated classification outcomes of decision tree classifiers and found the best decision tree model from them. In this work, CDT unpruned tree shows highest accuracy, precision,recall, f-measure, second highest AUROC and lowest RMSE than other models. Then, they extracted possible rules and significant features from this model and plasma glucose, plasma glucose 2hr after glucose and HDL-cholesterol have been found as the most significant features to predict the severity of Diabetes Mellitus [4].

S Alby, BL Shivakumar, tried to develop a method for the prediction of type 2 diabetes using adaptive neuro-fuzzy interface system (ANFIS) with genetic algorithms (GA). A comparative study has also been done with the result of our previous work in which General Regression Neural Network (GRNN) is applied. ANFIS with GA was applied for predicting diabetes disease and the most accurate result was obtained compared to our previous work in which we have tried the same using GRNN. The results strongly suggest that ANFIS with GA can aid in the field of diabetes disease. Optimization technique like Genetic Algorithm has much more influence on the accuracy of classification techniques. A combination of ANFIS and GA gave better results than the results achieved by using GRNN [5].

Tejas N. Joshi, Prof. Pramila M. Chawan, developed a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques. To predict diabetes via three different supervised machine learning methods including: SVM, Logistic regression, ANN. They propose an effective technique for earlier detection of the diabetes disease. The technique may also help researchers to develop an accurate and effective tool that will reach at the table of clinicians to help them make better decision about the disease status [6].

ChaithraNandMadhu B. analyses the various data mining techniques introduced in recent years to design a predictive model for cardiovascular diseases from the data obtained by transthoracic echo cardiographs.

This study investigates three different classification models: J48 Decision Tree, Naive Bayes and Neural Network on cardiovascular disease prediction and the same has been justified with the results of different experiments conducted and the performance of the models was evaluated using the standard metrics of Accuracy, Precision, Recall and F-measure [7].

ChandeepKaur and OlufemiMuibiOmisakin, evaluates different data mining methods capable of analyzing big clinical data. To extract information from clinical trial data mining methods to improve on them analytically, and above all to make a recommendation for the best possible data mining and analytical technique. This research was done as the number of cases of diabetes, especially Type 2 diabetes has increased recently as Type 2 diabetes accounts for nearly 90% of all cases of diabetes. This research evaluated a lot of data mining methods, but it is concluded that Big Data proves to be very efficient and effective as it can handle structured and unstructured data like medical records; diagnostic images etc. and interpret its data during the analysis step [8].

M.N. Sohail, RenJiadong, M.M. Uba, M. Irshad, Musavir Bilal, Usman Akbar, TahirRizwan examined machine learning classification and logistic regression modeling used to forecast the outcome of dataset through classification. They compared the classification analysis of previous researchers work by weka classification of experimenter part and auto-weka guide for better predictions. They establish a forecast prediction model on base of linear regression with the classification method by using weka performed on the real-life diabetes data set in order to acknowledge an education session for the medical specialists for the initial care of patients. Our results are based on the past researchers experience algorithms to perform the classification and we have pulled the relaxation accordance to free room of improvement in accuracy ratios and forecast analysis [9].

P. Suresh Kumar and V. Umatejaswi, tries to diagnose diabetes using k-means algorithm is used for gestational diabetes, cluster-1 for type-1 diabetes (juvenile diabetes), cluster-2 for type-2 diabetes. From this study, it was found out that data mining techniques can be used for predicting the type and risk levels of diabetes. Thus the proposed model helps in improving the diagnosis of the diseases which indeed helps in early cure of disease in the patients [10].

Mr. R. Sengamuthu, Mrs. R. Abirami, Mr. D. Karthik,explores the early prediction of diabetes using various data mining techniques. The analysis proves that Modified J48 Classifier provide the highest accuracy than other techniques. In the analysis of data mining techniques and tools Modified J48 Classifier gives 99.87% of highest accuracy using WEKA & MATLAB tool [11].

S.Sibi, S.Pushpalatha, provide a detailed survey on the use of data mining techniques in the field of diabetics. They also provide a survey on the recent and most useful research in diabetics using data mining techniques. Predicting diabetics using the risk factor also have certain drawbacks as the symptoms related to diabetics also is common with other diseases and so there is a need for a better data mining model that can effectively identify the occurrence of diabetics. This can be done by building an effective model that takes into consideration all the factors related to diabetics disorder. This survey provides a path for future research in application of data mining in diagnosis of diabetics [12].

D. Jeevanandhini , E. Gokul Raj ,V. Dinesh Kumar, N. Sasipriyaa, compared the performance of the algorithm using the data mining techniques. K-means clustering algorithm and Classification algorithm like SVM , KNN, J48, Random forest. They obtained that SVM has higher accuracy comparing to other three algorithms [13].

IV. Methodologies

Figure 1 represents framework of working methodology. The methodology work starts with problem definition followed by data collection and pre-processing steps. The pre-processing part includes data selection and data transformation. After pre-processing it precedes with classification techniques with pruning which finally leads to discovering knowledge which is to be benefitted. [22]

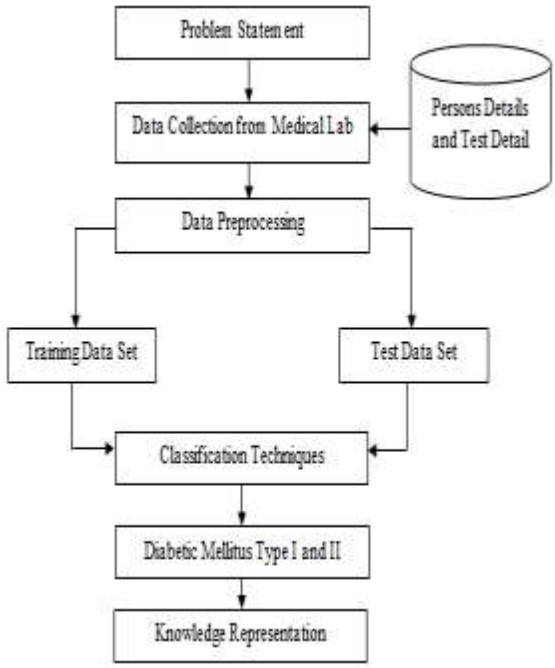


Figure 1 Data mining work methodology [22]

In order to research the high-risk group of Data Mining, we need to utilize advanced information technology. Therefore, data mining, also known as knowledge Discovery Database (KDD), is defined as the computational process of discovering patterns in large datasets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database system.

Table 4.1: Descriptive Statics

Algorithm	Use	Formula	Result
Logistic Regression	LR is useful for situations in which can be able to predict the presence or absence of a characteristic or outcome based on values of set of predictor variables.	$H(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_pY_p)}}$	-1.7122 * BPAD + 1.3842 * BWAN + -1.1221 * BWAN2 + 0.0109 * BMAN2 + 0.5552
GAUSSIAN PROCESS CLASSIFIER	It is a collection of random variables, any infinite number of which have a joint Gaussian distribution [6].The Gaussian distribution is over vector, whereas the Gaussian process is over functions.	$k(x,x')=\theta_0\exp\{-\frac{1}{2}\sum_{l=1}^{d}w_l(x_l-x_l')^2\}+\theta_1$	Inverted Covariance Matrix * Target-value Vector: Lowest Value = -0.6409469815752027 Highest Value = 0.37328157457484096
Multilayer Perception	Multilayer Perceptron (MLP) is gained their popularity due to it is a simple architecture but a powerful problem-solving ability.	$E(\vec{w};D)=\frac{1}{2}\sum_{i=1}^{P}\|\|y(\vec{x}^{(i)};\vec{w})-\vec{d}^{(i)}\|\|^2$	Threshold and Node 5 grouping dataset
Simple Linear Regression	It is a model that assumes a linear relationship between the input variables (x) and the single output variable (y). When there is a single input variable (x), the method is referred to as **simple linear regression**.	$y_i=\alpha+\beta*x_i+\varepsilon_i$	Linear regression on BWAN 0.33 * BWAN + 0.56

SMOreg	SMO implements the sequential minimal optimization algorithm for training a support vector classifier using polynomial or Gaussian kernels (Platt, J. 1998 and S. Keerthiet. al., 2001). Missing values are replaced globally, nominal attributes are transformed into binary ones, and attributes are normalized by default.	Predicted price = + 0.2209 * (normalized) SEX - 0.3164 * (normalized) GRADE + 0.3937	Number of kernel evaluations: 5050 (87.422% cached)

V. Results and Analysis

Authors have used five different algorithms: linear regression model, Gaussian Process, Multilayer Perception, Simple Linear Regression and SMOreg. These processes were selected for inclusions in this study because of their popularity in the recently published literature as wellas their good performance even in fewer amounts of training data. Below given is a short description about the selected algorithms for study.

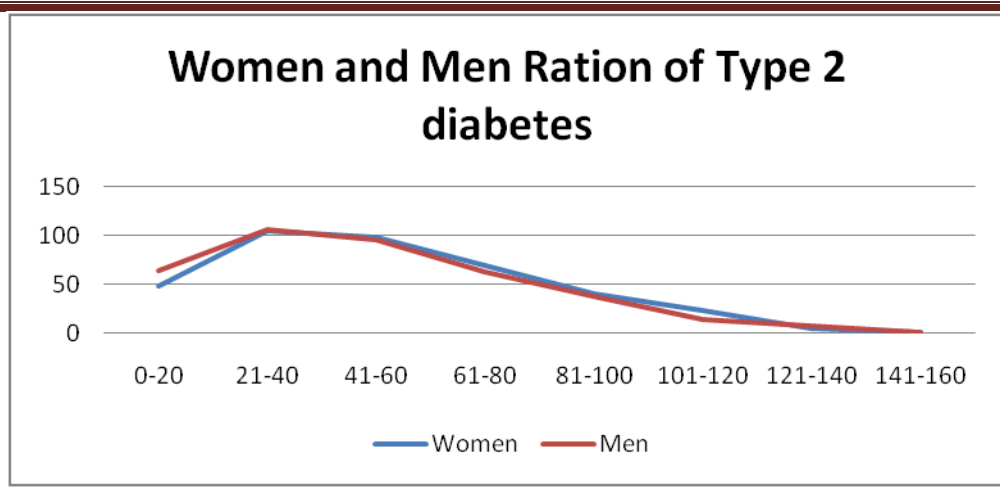
We used different data mining classification algorithms for making predictions in the domain of medical diagnosis and analyze their efficiency. For analyzing the efficiency ofthe system, the total data set collected was divided into two parts viz training data and test data [15].

Authors takeMetadata Created on March 30, 2017 and Metadata Updated Date on April 11, 2018 from Allegheny County / City of Pittsburgh / Western PA Regional Data Center publisher and maintain by Lynda Jones. Source Schema Version of dataset has 1.1 version in Health Category.

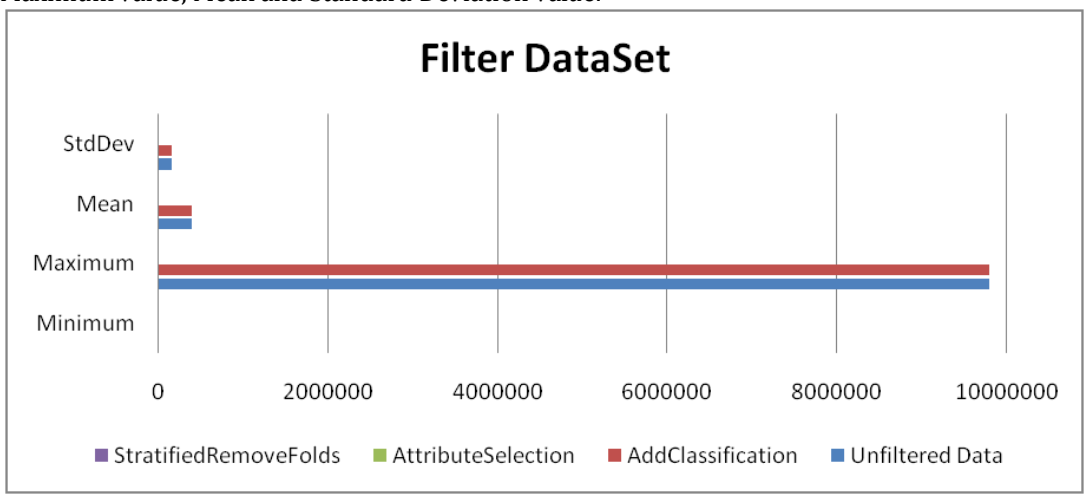
Table 5.1 Used dataset Categories

Code	Description
CT	Census Tract
BPAD	Total number of members, who are Allegheny County residents and who were effectively enrolled for at least 90 continuous days during the 2015 calendar year (Jan 1 - Dec 31).
BPAN	(Diagnosed): Total number of members who met the denominator criteria and are diagnosed with Type 2 diabetes mellitus.
BPAN 2	(Diagnosed & Medicated): Total number of members who met the denominator criteria, who are diagnosed with Type 2 diabetes mellitus, and had at least one Type 2 diabetes related medication claim.
BWAD	Total number of women members, who are Allegheny County residents and who were effectively enrolled for at least 90 continuous days during the 2015 calendar year (Jan 1 - Dec 31).
BWAN	(Diagnosed): Total number of women members who met the denominator criteria and are diagnosed with Type 2 diabetes mellitus.
BWAN 2	(Diagnosed & Medicated): Total number of women members who met the denominator criteria, who are diagnosed with Type 2 diabetes mellitus, and had at least one Type 2 diabetes related medication claim.
BMAD	Total number of male members, who are Allegheny County residents and who were effectively enrolled for at least 90 continuous days during the 2015 calendar year (Jan 1 - Dec 31).
BMAN	(Diagnosed): Total number of male members who met the denominator criteria and are diagnosed with Type 2 diabetes mellitus.
BMAN 2	(Diagnosed & Medicated): Total number of male members who met the denominator criteria, who are diagnosed with Type 2 diabetes mellitus, and had at least one Type 2 diabetes related medication claim.

Table 5.1 was identifying the dataset columns which can identify all the values of data description about diabetes medication. These categories of data helps to analyses and Determination of health care diseaselikediabetes medication. On Basis of these dataset, we made one graphical representation, to identify the different ration of women and male category diagnosed with Type 2 diabetes mellitus.



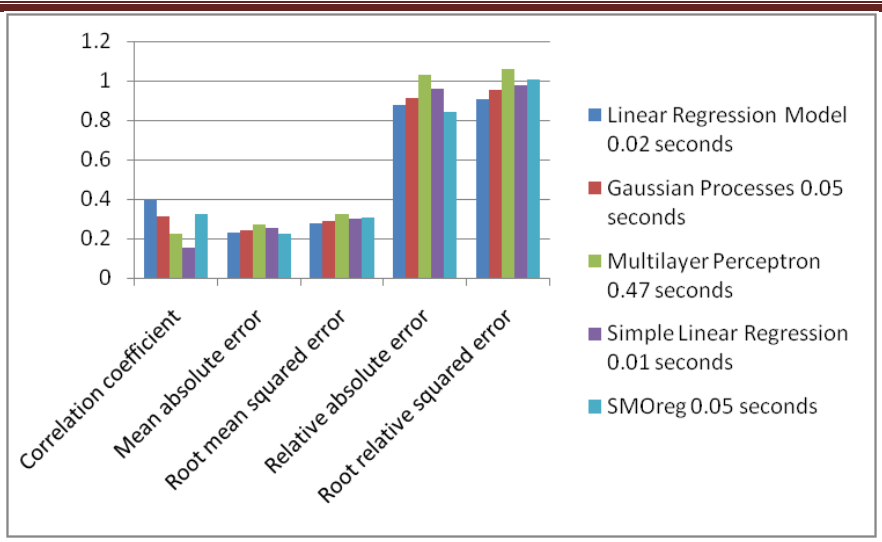
These influencing factors are categorized as input variables. For this work, recent real world data is collected from Allegheny County / City of Pittsburgh / Western PA Regional Data Center. This data is then filtered out using manual techniques like Stratified Remove Folds, Attribute Selection, Add Classification and Unfiltered Data. Authors analyzed this filter to identifying different values with its parameters like Minimum value, Maximum value, Mean and Standard Deviation value.



Authors apply function type classification data mining techniques like Linear Regression Model, Gaussian Processes, Multilayer Perceptron, Simple Linear Regression and SMOreg with 100 Instances, 10 attributes with 10-fold cross-validation in Test mode.

Table 5.2: Descriptive Statics

Classifier function	Time taken to build model	Correlation coefficient	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
Linear Regression Model	0.02 seconds	0.3957	0.231	0.2764	87.927 %	90.5877 %
Gaussian Processes	0.05 seconds	0.3126	0.2394	0.2899	91.122 %	95.0031 %
Multilayer Perceptron	0.47 seconds	0.2207	0.2711	0.3229	103.1621 %	105.8187 %
Simple Linear Regression	0.01 seconds	0.1545	0.2514	0.2984	95.665 %	97.7956 %
SMOreg	0.05 seconds	0.3216	0.2214	0.3069	84.2451 %	100.5754 %



We can identify from this algorithm with different categories.Simple Linear Regression is takenTime to build model in 0.01 seconds, Correlation coefficient with low values 0.1545, Mean absolute error in 0.2514, Root mean squared error in 0.2984, Relative absolute error in 95.665% and Root relative squared error in 97.7956 %. So from this analysis we analyzed that Simple Linear Regression generates the best result for this type of Dataset.

The results suggest that these algorithms can performgood prediction with least error in the field of medicaldiagnosis and can be used to build various tools for prognosisand thus can supplement medical doctors in performingexpert diagnosis. A limitation of this study is thatthe data set under consideration limits to physiologicalparameters only, which can be increased to include clinicalparameters also, thus the comparative study of predictionusing clinical parameters can be carried out [15].

VI. Conclusion and Future Task

In this paper, function type classification techniques are used for prediction on thetraining dataset of 390 values, to predict andanalyzed disease's performance. In this study, basedon selected heath care related input variables collected from real world. Among all datamining classifiers Simple Linear Regression performs best with High accuracy and therefore SLR proves to bepotentially effective and efficient classifier algorithm. Also comparison of all 5 classifiers with the help of WEKA experiments is also performed, in this case also SLR proves to be the best in higher than other classifiers.Forthe future research work, it is suggested that fordecision rule generation roughset theory can also beused to handle noisy, missing and uncertainty data.

References

1. DeeptiSisodiaa, Dilip Singh Sisodia, Prediction of Diabetes using Classification Algorithms , International Conference on Computational Intelligence and Data Science (ICCIDS 2018) Procedia Computer Science 132 (2018) 1578–1585
2. Rafiqul Islam, ObaidurRahman, The Risk Factors of Type 2 Diabetic Patients Attending Rajshahi Diabetes Association, Rajshahi, Bangladesh and Its Primary Prevention, Food and Public Health 2012, 2(2): 5-11
3. S.Ummugulthum and Natchiar, S.Baulkani, REVIEW OF DIABETES DISEASE DIAGNOSIS USING DATA MINING AND SOFT COMPUTING TECHNIQUES, International Journal of Pure and Applied Mathematics Volume 118 No. 10 2018, 137-142ISSN: 1311-8080 ,ISSN: 1314-3395
4. Koushik Chandra Howlader_, Md. ShahriareSatu y, AvijitBarua _, Mohammad Ali Moni, Mining Significant Features of Diabetes Mellitus Applying Decision Trees: A Case Study InBangladesh, bioRxiv preprint first posted online Nov. 30, 2018.
5. S Alby, BL Shivakumar, A prediction model for type 2 diabetes using adaptive neuro-fuzzy interface System, Biomedical Research 2018; Special Issue: S69-S74, ISSN 0970-938X
6. Tejas N. Joshi, Prof. Pramila M. Chawan,Dewangan.et.al. Int. Journal of Engineering Research and Application www.ijera.com ISSN: 2248-9622, Vol. 8, Issue 1, (Part -II) January 2018, pp.-09-13
7. Chaithra N and Madhu B, Classification Models on Cardiovascular Disease Prediction using Data Mining Techniques, Journal of Cardiovascular Diseases & Diagnosis, Chaithra and Madhu, J Cardiovasc Dis Diagn 2018,6:6
8. ChandeepKaur and OlufemiMuibiOmisakin, Data Mining Methods to Improve Clinical Trials in Diabetic Patients, iMedPub Journals, Annals of Clinical and Laboratory Research, ISSN 2386-5180, Vol.6 No.4:266 2018

9. M.N. Sohail, RenJiadong, M.M. Uba, M. Irshad, Musavir Bilal, Usman Akbar, TahirRizwan, Forecast Regression analysis for Diabetes Growth: An inclusive data mining approach, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 9, September 2018, ISSN: 2278 – 1323
10. P. Suresh Kumar and V. Umatejaswi,Diagnosing Diabetes using Data Mining Techniques, International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017 ISSN 2250-3153
11. Mr. R. Sengamuthu, Mrs. R. Abirami 2, Mr. D. Karthik,VARIOUS DATA MINING TECHNIQUES ANALYSIS TO PREDICT DIABETES MELLITUS,International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072
12. S.Sibi1, S.Pushpalatha, Survey on Data Mining Techniques for Diabetics Research and Analysis, International Journal of Innovative Research in Science, Engineering and Technology An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 1, March 2018
13. D. Jeevanandhini , E. Gokul Raj ,V. Dinesh Kumar, N. Sasipriyaa, Prediction of Type2 Diabetes Mellitus Based on Data Mining, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181
14. GanjarAlfian , Muhammad Syafrudin, Muhammad FazalIjaz , M. Alex Syaekhoni,Norma LatifFitriyani and Jongtae Rhee, A Personalized Healthcare Monitoring System forDiabetic Patients by Utilizing BLE-Based Sensors and Real-Time Data Processing, Sensors 2018, 18, 2183.
15. AbidSarwar and Vinod Sharma, Comparative analysis of machine learning techniques in prognosis of type II diabetes, ISSN 0951-5666, Volume 29, Number 1, 16 April 2013, pp. 123–129.
16. MS.TEJASHRI N. GIRI and PROF. S.R.TODAMAL, DATA MINING APPROACH FOR DIAGNOSING TYPE 2 DIABETES, ISSN: 2348-4098, VOL 2 ISSUE 8 NOV-DEC 2014, pp. 191-194.
17. Alberti, K. G., and Zimmet, P. Z., Definition, diagnosis and classification of diabetes mellitus and its complications, part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet. Med. 15(7):539–553, 1998
18. Dr. T. Karthikeyan and K.Vembandadsamy, An Analytical Study on Early Diagnosis and Classification of Diabetes Mellitus, International Journal of Computer Application, Volume 5– No. 5, August 201, 96-104.
19. S.Ummugulthumnatchiar And S.Baulkani, Review Of Diabetes Disease Diagnosis Using Data Mining And Soft Computing Techniques, International Journal Of Pure And Applied Mathematics, Issn: 1314-3395, Volume 118 No. 10 2018, Pp. 137-142.
20. Prof. Samiksha H. Zaveri And Dr. Narayan Joshi, A Comparative Study Of Data Analysis Techniques In The Domain Of Medicative Care For Disease Predication, International Journal Of Advanced Research In Computer Science, Issn No. 0976-5697 , Volume 8, No. 3, March – April 2017, Pp. 564-566.
21. Kawsar Ahmed And TasnubaJesmin, Comparative Analysis Of Data Mining Classification Algorithms In Type-2 Diabetes Prediction Data Using Weka Approach, International Journal Of Science And Engineering, Issn: 2086-5023, Vol. 7(2), Pp.155-160, October 2014.
22. T.Nithyapriya And S.Dhinakaran, Analysis Of Various Data Mining Classification Techniques To Predict Diabetes Mellitus, International Journal Of Engineering Development And Research, Volume 5, Issue 4, Issn: 2321-9939, Pp. 695-703.
23. Miss. N. Vijayalakshmi And Miss. T. Jenifer, An Analysis Of RiskFactors For Diabetes Using Data Mining Approach, International Journal Of Computer Science And Mobile Computing, Issn 2320–088x, Vol.6 Issue.7, July- 2017, Pg. 166-172.
24. Chang, Huan-Cheng; Chang, Pin-Hsiang; Tseng, Sung-Chin; Chang, Chi- Chang; Lu And Yen-Chiao, A Comparative Analysis Of Data Mining Techniques For Prediction Of Postprandial Blood Glucose: A Cohort Study, International Journal Of Management, Economics And Social Sciences Special Issue-International Conference On Medical And Health Informatics, Issn 2304 – 1366, Vol. 7(S1), Pp.132 – 141.
25. Boyu Wang, Feng Wan, Peng Un Mak, Pui In Mak, And Mang I Vai, Eeg Signals Classification For Brain Computer Interfaces Based On Gaussian Process Classifier, Ieee, Doi: 10.1109/Icics.2009.5397570 · Source: IeeeXplore, January 2010.
26. Http://Ml.Informatik.Uni- Freiburg.De/Former/_Media/Documents/Teaching/Ss09/Ml/Mlps.Pdf
27. Https://Machinelearningmastery.Com/Linear-Regression-For-Machine-Learning/
28. Https://Www.Researchgate.Net/Post/How_Do_I_Use_The_Smoreg_Model_Equation_So_That_When_I_Enter_New_Data_It_Should_Give_Me_The_Predicted_Price