Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 5, 8588-8601

Prediction Using Machine Learning Algorithms For Type 2 Diabetes Mellitus

Kritika Sinha[1] , Sunita Kushwaha[2] , Varsha Thakur[3]

[1]Research Scholar, MATS School of Information Technology, MATS University, Raipur (C.G.)
kkritikasinha@gmail.com
[2]Assistant ProfessorAssistant Professor, MATS School of Information Technology, MATS University, Raipur (C.G.)
drsunitak@matsuniversity.ac.in
[3]Assistant Professor, Govt. NPG Science College, Raipur (C.G.), varshathakur1308@gmail.com

ABSTRACT

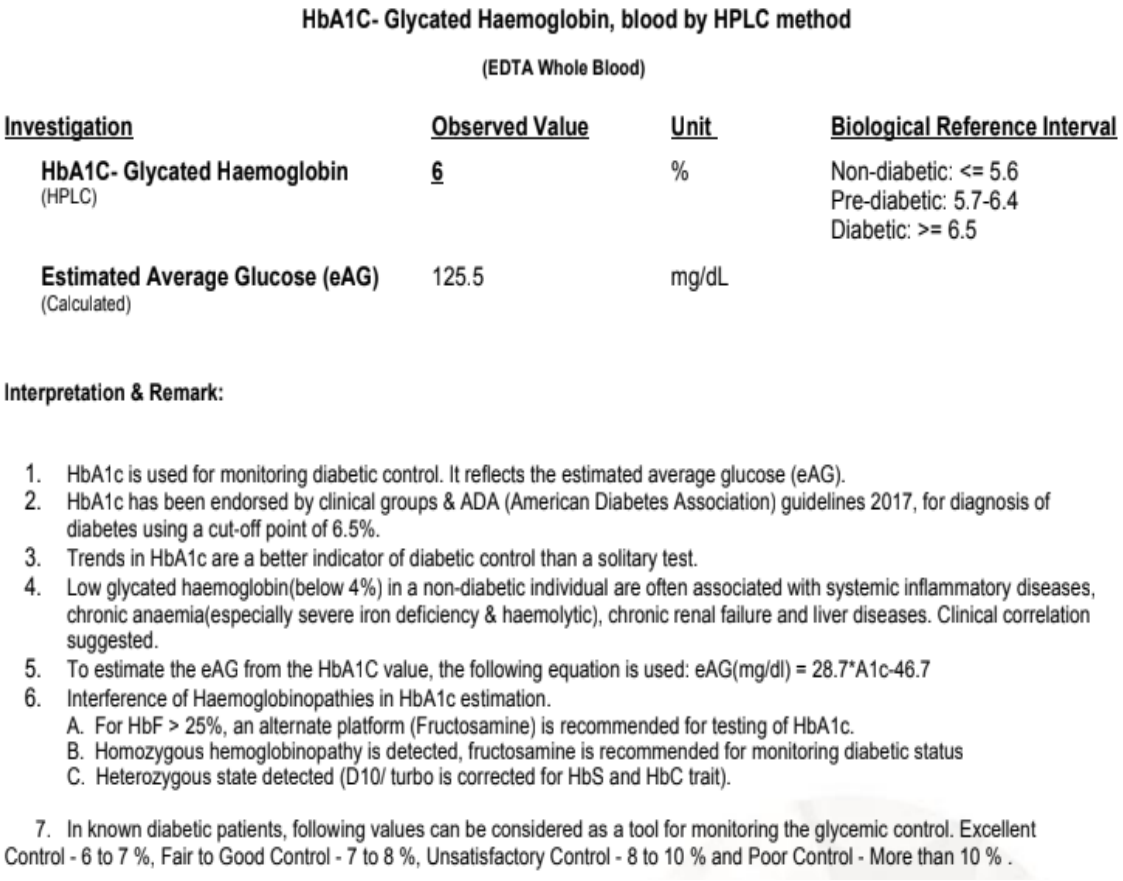
Diabetes is precarious health issue and huge population of India are afflict from it. Entire world is adversely affected by this problem. In the modern world, it affects the any individual regardless of Age, the factor leading to diabetes problem are fatness, living style, bad diet, high blood pressure, less physical activity, etc. People suffering from diabetes have more chance of getting stirred of various diseases like stroke, eye problem, heart disease, kidney disease, nerve damage, etc. Data analysis concepts are helpful in detection of complication of diabetes at the primary stage and prevent the patient form the bed effects of diabetics. Healthcare industries generate huge amount of data which is used for analysis. Diabetes must be prevented and cured in order to enhance the lives of all those who are impacted by it. Data analysis concepts are helpful in the detection and prevention of the complication of diabetes at the primary phase. This paper studies the diabetes data of various state of India. According to the data obtained, prevalence of diabetes in percentage is almost half in rural area as compare to urban areas; prevalence of pre-diabetes is approximately 10% to 20% less in rural area than urban areas. The experimental observation shows that the performance of random forest and SMO are surpass than logistic regression, naive base and decision tree. The accuracy of random forest is highly acceptable than others.

KEYWORDS: Logistic Regression, Decision tree, Naive Bayes, SMO.

1. INTRODUCTION

Diabetes is the prevailing medical conditions buzzed nowadays. It precipitate crippled as well as death in some situations. Diabetes mellitus (DM) is a congenital anomaly that is signalized by high blood glucose. People suffering from diabetes have risk of some other chronic diseases related to heart and kidney. In India approximately 31.7 million people affected by diabetes in 2002 which is the highest in the list of entire countries in world also predictions says that diabetes mellitus may affects approximately up to 79.4 million people of India till 2030. This is the dangerous ituation in India because pervasiveness of diabetes is higher than western countries. It signifying that body mass index (BMI) related to diabetes is lower in Indians compared with Europeans. Therefore, relatively lean Indian adults with a lower BMI may be at equal risk as those who are obese. As Indians has more predisposal to development of the chronicle disease like diabetes at an early age which indicate that screening of diabetes must be carefully done despite of patient agein India.(**Seema Abhijeetet.al, 2019)(Pradeepa R. et. al, 2021)**.Pradeep R. and others represent the data weighted prevalence of diabetes and pre-diabetes for ruler and urban areas of the states of India. Prevalence of diabetes in percentage is almost half in rural area as compare to urban areas and Prevalence of pre-diabetes is approximately

10 to 20% less in rural area than urban areas **(Pradeepa R. et. al, 2021)**. Glycated hemoglobin (HbA1C) value is used in the detection of diabetes and pre diabetes condition, in the human when this value is <=5.6 then its shows Non diabetes, when it is between5.7 to 6.4 this value is denote pre- diabetes stage, and value >= 6.5 is denotes diabetes condition.



HbA1C- Glycated Haemoglobin, blood by HPLC method

(EDTA Whole Blood)

Investigation	Observed Value	Unit	Biological Reference Interval
HbA1C- Glycated Haemoglobin (HPLC)	6	%	Non-diabetic: <= 5.6 Pre-diabetic: 5.7-6.4 Diabetic: >= 6.5
Estimated Average Glucose (eAG) (Calculated)	125.5	mg/dL	

Interpretation & Remark:

1. HbA1c is used for monitoring diabetic control. It reflects the estimated average glucose (eAG).
2. HbA1c has been endorsed by clinical groups & ADA (American Diabetes Association) guidelines 2017, for diagnosis of diabetes using a cut-off point of 6.5%.
3. Trends in HbA1c are a better indicator of diabetic control than a solitary test.
4. Low glycated haemoglobin(below 4%) in a non-diabetic individual are often associated with systemic inflammatory diseases, chronic anaemia(especially severe iron deficiency & haemolytic), chronic renal failure and liver diseases. Clinical correlation suggested.
5. To estimate the eAG from the HbA1C value, the following equation is used: eAG(mg/dl) = 28.7*A1c-46.7
6. Interference of Haemoglobinopathies in HbA1c estimation.
   A. For HbF > 25%, an alternate platform (Fructosamine) is recommended for testing of HbA1c.
   B. Homozygous hemoglobinopathy is detected, fructosamine is recommended for monitoring diabetic status
   C. Heterozygous state detected (D10/ turbo is corrected for HbS and HbC trait).
7. In known diabetic patients, following values can be considered as a tool for monitoring the glycemic control. Excellent Control - 6 to 7 %, Fair to Good Control - 7 to 8 %, Unsatisfactory Control - 8 to 10 % and Poor Control - More than 10 % .

Fig 1: HbA1C Report screen short

The growing rate of this chronicle disease is adding various other health issues as a risk factor. Various researches focused on that problem and conducting analytical study for the detection of health related issues with the help of machine learning algorithms. This study observed that machine-learning algorithms such as clustering, association and classification, works better in detection of different health problem and diseases**. (Sisodiya D. etal.,2018) (Mazumdar A. etal.,2019).** From Fig 2 and Fig 3 it is observed that the diabetes mellitus affected the more population of urban areas with respect to the rural areas but the pre diabetes condition is lightly varied in the rural and urban population as the life style and food habits of urban and rural area is quite different to each other, its play a imperative role in the diabetes condition of India. Also the conversion rate of pre-diabetes to diabetes stage is higher in urban areas as compare with rural area.

Table 1: Weighted prevalence of diabetes and pre-diabetes in 15 states /union territory of India- the ICMR INDIA study (Pradeepa R. et. al, 2021).

States/Union Territory	Prevalence of diabetes (%)	Prevalence of pre-diabetes (%)

	Urban	Rural	Overall	Urban	Rural	Overall
Andhra Pradesh	12.6	6.3	8.4	11.1	9.6	10.1
Arunachal Pradesh	5.8	4.9	5.10	14.2	12.3	12.8
Assam	12.4	4.4	5.5	13.6	11.6	11.9
Bihar	10.5	3.5	4.3	15.5	9.3	10.0
Chandigarh	14.2	8.3	13.6	14.5	14.7	14.6
Gujrat	9.5	5.1	7.1	8.4	11.5	10.2
Jharkhand	13.5	3.0	5.3	10.7	7.4	8.1
Karnataka	11.1	5.6	7.7	14.1	10.2	11.7
Maharashtra	10.9	6.5	8.4	15.2	11.1	12.8
Manipur	7.1	4.4	5.1	7.2	7.5	7.5
Meghalaya	8.9	3.5	4.5	7.4	10.6	10.0
Mizoram	7.9	3.6	5.8	6.2	5.8	6.0
Punjab	12.0	8.7	10.0	8.6	7.9	8.2
Tamil Nadu	13.7	7.8	10.4	9.8	7.1	8.3
Tripura	15.5	7.2	9.4	16.2	14.2	14.7

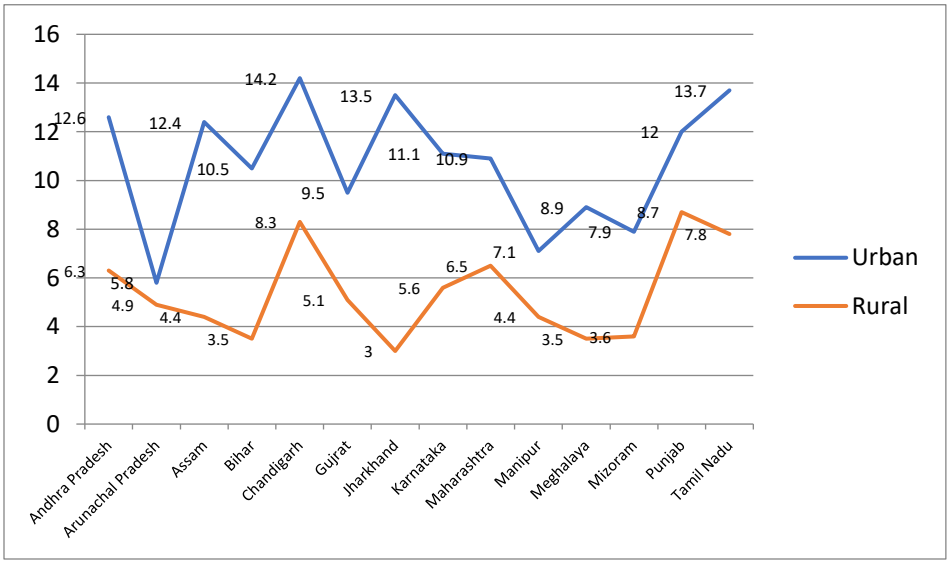


Fig 2: Prevalence of Diabetes in Rural and Urban Areas

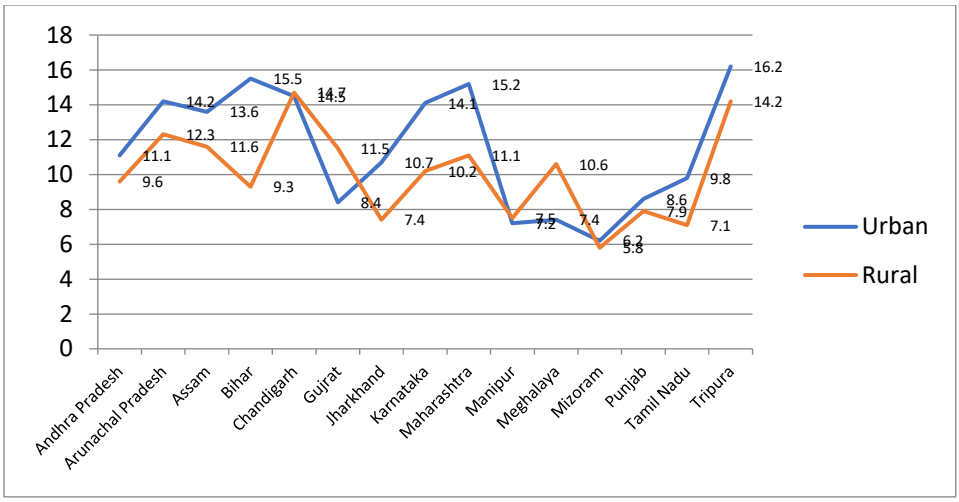


Fig 3: Prevalence of Per-Diabetes in Rural and Urban Areas

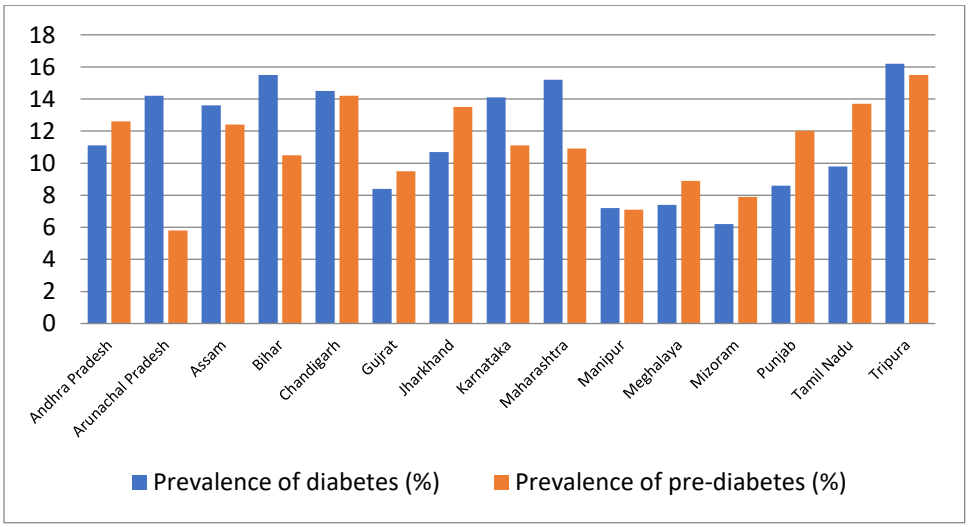


Fig 4: Prevalence of Diabetes and Per-Diabetes in Urban Areas

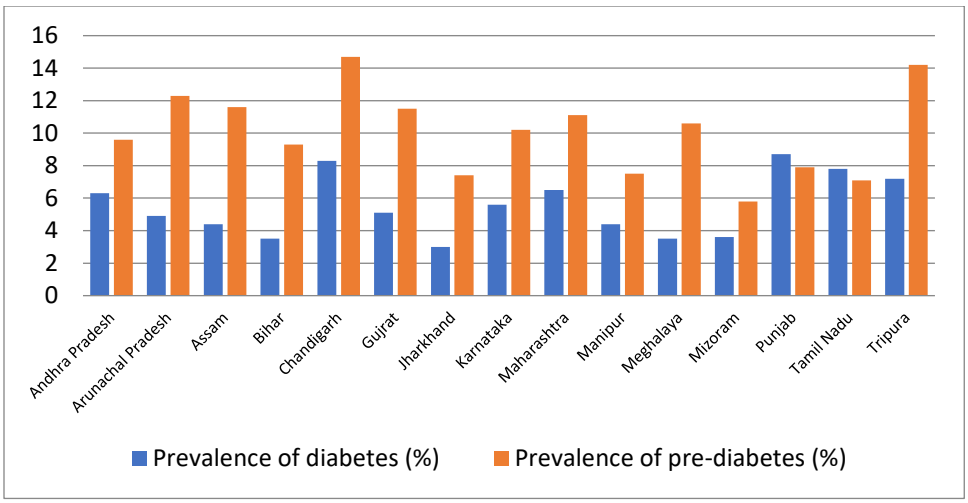


Fig 5: Prevalence of Diabetes and Per-Diabetes in Rural Areas

2. DATA MINING

In machine learning, Data mining is prominent concepts, it extracts information from a given data set and it digonsis trends, patterns, and rules. Data mining works along with predictive analysis, The predictive analysis identifies patterns in huge amounts of data. Data mining useds to recognize patterns in datasets for problem associated with domain specific.**(Sneha N. etal., 2019)(Aravindan J. et al., 2019).**

Data Mining Techniques catogarized in two parts supervised and unsupervised, Supervised Data mining also considered as Predictive mining-Supervised learning is the Data mining task of inferring a function from labeled training data. In supervised learning output value label specified in early stage and data is divided in two parts first is training dataset and second is testing data set. In supervised learning, each record has an input (typically a vector) and a desired output value. Second categories of data mining is Unsupervised Data Mining known as Descriptive-unsupervised searches hidden facts or pattern from unlabeled data after that it provides label with respect to the similarty(dissimilarty).

2.1. Supervised data mining (Predictive)

supervised learning basically considers prediction, classification and regression.

Prediction: Prediction searches hidden pattern from current or historical data . It provide a look into what are the current trends or needs.

Classification- Classification analyzed attributes, along with their association .Classification mainly deal with discrete value.

Regression- Regression identifies hidden relation among variables in a dataset.. Regression is a straight forward white box technique. forecasting and data modeling uses regression techniques (**Aishwarya M. et al., 2019) (Sisodiya D. et al., 2018)(Azzar A. et al., 2018)**

2.2. Unsupervised data Mining (Descriptive)

unsupervised learning related to association, clustering and sequential patterns

Association-Association techniques are puffed up in market basket analysis. But it is popular data mining technique and used in a range of statistics analysis. It indicates how some data relats to other data example. For example Purchase of hamburgers is frequently accompanied by that of french fries.

Clustering- Clustering is grouping datasets with similar behaviour. Similar group should have similar properties and different group should have dissimilar properties. Clustering is deal with unlabelled dataset and it provides a different ClusterID or Cluster_name to the dataset

Sequential patterns- Sequential pattern mining is used to uncover the sequential patterns or series of events. It is generally acknowledged discrete values; therefore temporal data mining concepts are related to it. It is often beneficial in shelf placement and promotions. (**Aishwarya M. et al., 2019) (Sisodiya D. et al., 2018).**

3. DIABETES MELLITUS

Diabetes mellitus is a precarious disease which causes high blood sugar. Insulin harmones transfers sugar from the blood into body cells which is used for synthesis of energy. In diabetes the normal release and use of insulin is getting affected, this abnormal change in insulin is triggers this disease**(Wu H. et al. 2017). W**hen blood glucose suit is high from normal condition than diabetes is detected. Our body gets this glucose form our foods **(Pattekari S.A. et al., 2018)**

Diabetic Symptoms and its Effects - The general symptoms of diabetes are increased in appetite, feeling thrusty, Excessive Weight loss, Frequent Urination, Blurry invision and Extreme Fatigue. Diabetes can effects different parts of the body such as Risk of Stroke, Loss of consciousness, Extreme thrust, disturbance in visualization, Sweet smelling breathes,Contracts and Glaucoma, Risk of heart diseases, Risk of infections, Kidney Damage, High Blood Pressure,Gastro paresis and Excessive urination.

4. LITERATUREREVIEW

As the life style of present generation intend to epic center of various health problem, therefore study of health sector is strengthen rapidly to provide better solution. In the form of reports huge amount of data were collected from the last few decades. Hence the machine learning approaches come to the picture as it become possible to extract curtail information, facts and pattern through the data collected in early time periods. Data analysis of health sector data is frequently increased and predictive data mining is used for disease diagnosis at the primary phase it helps physicians to cure patient form the major affects. The available literature reveals important facts, which are related to the diabetes and role of data mining approached in diabetes studies. At the present time various algorithm and techniques must be used for data analysis such as Decision tree, SVM, KNN, classification etc. **(A. Singh et al. , 2017) (Pattekari S.A. et al., 2012) (Zhe W. et al.,2015)** review is organized in the table2.

Table 2: Summery of Research Work in the field of Data analysis for diabetes patient

S.No NCE NO.	AUTHOR NAME	PROBLEM	ALGORITHM USED	TOOL USED	CONCLUSION
1	R.K. Kavitha and W. Jai Singh **(Kavitha R.K. et al., 2020)**	They are present a better prediction for risk related to diabetes using 3 algorithm Decision, Tree Multilayer Prceptron and Naïve Bayes with 5 or 8 k-fold values.	Decision tree, naïve bayes algorithm, multi layer perception	NA	They proposed a model helps to detect early diabetes among patients with high accuracy.
2	Kishore N.G., Rajesh V., Vamsi A., Reddy A, Sumedh K., Reddy TRS (**Kishore N.G., et. al, 2020)**	They present a comparative study among 5 well known machine learning algorithms	SVM, KNN, logistic regression, random forest	NA	They concluded that on their dataset random forest perform better than others.
3	T.M. Allam, M. A. Iqbal, A.Yasir, A. Wahab,S. Ijaz, T.I. Baig **(Allam T.M, et al., 2019)**	A few existing classification methods for medical diagnosis of diabetes patients have been discussed on the basis of accuracy	Apriori, random forest, ANN, K-mean	NA	A classification problem has been detected in the expressions of accuracy. diabetes very much related to the glucose level BMI this relation extract by apriori, and other are used in prediction.
4	N.Sneha, T.Gangil **(Sneha N., et al., 2019)**	The objective of this research is to make use of significant features selection for better predication using Machine learning and find the optimal classifier to give the closest result comparing to clinical outcomes.	Data Mining, Random forest and naïve bayes	NA	According to their experiment random forest perform better than other.

5	J. Aravindan **(Aravindan J. et al., 2019)**	They used checking and monitoring of family members of diabetes type 2 patients, and this may help the screening of early detection or pre diabetes detection of risk factors in them. Hazard like obesity, FH, consanguinity etc. are discussed.	Diabetes mellitus; Young onset diabetes, chi-square test.	NA	They used checking and monitoring of family members of diabetes type 2 patients, and this may help the screening of early detection or pre diabetes detection of risk factors in them.
6	Deepti sisodia, dilip singh sisodiya **(Sisodia D. et al., 2018)**	They used Pima India dataset for diabetes and aimed to design more accurate model for prediction. They select 3 algorithms for evaluation namely Decision tree , SVM, Naïve bayes algorithm	Decision tree , SVM, Naïve bayes algorithm	weka	Naive Bayes outperforms with the highest accuracy of 76.30% comparatively other algorithms.
7	Nazim Razali ,Aida Mustapha , Syed Zulkarnain Syed Idrus , Mohd Helmy Abd Wahab , Siti Aida Fatimah Madon **(Nazim R.. et al. , 2019)**	This paper present the study of some well known classification algorithm and check their performance against some evaluation parameters such as recall, precision etc.	Naive Bayes, Sequential Minimal Optimizatio n (SMO), Decision Tree and Simple Logistic Regression	NA	With the help of confusion matrix performance of selected algorithms are evaluated with respect to the precision, recall etc.
8	Ramachandra majji, Bhramaramb aravi **(Ramachand ro M. et al., 2018)**	To create a risk prediction website for diabetes using the survey data.	FP tree classifier algorithm	Wek a	The website help to predict that a particular person getting affected by diabetes in future. It is questioner based prediction site.
9	M.S. Kadam, I.W.Ghindaw i, D.E. Mhawai	They proposed a classification algorithm which uses the K-nearest method and for remove the unwanted	KNN,Deciss ion Tree.	NA	The new algorithm use Decision Tree and class level of sample data. Experiments, shows

	(M.S. Kadam et al., 2018)	data. This approach helpful to reduce the processing time.			that the proposed algorithm gives 98.7% accurate result.
10	Han wu shengiyang, z hangin huang **(Wu H. et al. 2017)**	They attempt to increase the accuracy of prediction of model and build a more adapted model. Which perform well for different data set.	Logistic Regression, K-means cluster algorithm	weka	They present new model and compare with k mean, logistic regression. Results shows that proposed one is 3.04% more accurate.
11	Quqn Zou, Kaiyang Qu, Yameilua **(Q. Zou et al., 2018)**	In this study, five-fold cross validation was used to examine the models..	Random forest, decision tree, neural network	NA	Comparing the results of three classifications methods random forest, decision tree and neural network, random forests are better than the another classifiers methods
12	Deeraj shetty, Kishor Rit, Sohail Shaikh, Nikita Patel **(Shetty D. et al. ,2017).**	Naïve Bays and KNN algorithms are used for the prediction of diabetes dataset. And developed a software for prediction.	KNN and Naive Bayes Algorithm	NA	They developed an expert software for prediction of diabetes problem using the patient record as input.
13	Asmita singh,Malha .N. Halgamuge,Rajaseharam Lakshmigant han (**A. Singh et al. , 2017)**	They evaluate the effects of different data types (Numeric or text only) on some algorithms namely Naïve bayes Random Forest, K-nearest Neighbour Algorithm	Naïve bayesRandomForest,K-nearest Neighbour Algorithm	NA	Random forest dose not affected by changing number of data. Naïve bayed is good for independent feature variables in problem space. And RF and KNN provide more accuracy than Naïve bayes.
14	Shadab adam Pattekari,Asma Parveen **(Pattekari S.A. et al., 2012) (Zhe W. et al., 2015)**	They developed a web based intelligent system for prediction using naïve bayes algorithm	Naive Bayes algorithm	NA	The system use historical dataset related to heart disease. And provide better system for prediction of heart problem and their available solution .

15	Wei Zhe, Ye Guangjian **(Zhe W. et al., 2015)**	They used data collected from tertiary referral hospital in Lanzhou for the period of 2009, January to 2014, march. They analysis this data with the help of FP-tree, Apriori and propose Improve FP-tree.	Apriori Algorithm, FP-tree Algorithm	C# program	They propose a new and modified algorithm, improved FP-tree for prediction of diabetes data. The newly proposed algorithm provide better result.
16	V.Anuja kumara,R.chitra **(Kumari V.A. et al., 2013)**	They present a study on diabetes dataset from the university of California , Irvine, Using Support Vector Machine (SVM).	Support vector machine	NA	On the basis of the experiments of this paper SVM can perform better in diabetes detection.

(Kritika S. et al., 2022)

5. METHODOLOGY

Data mining algorithm play important role in the prediction of future output on the basis of already available information. The paper examine the pima India dataset for diabetes problem analysis with the help of some well known data mining algorithm namely Sequential Minimal optimization (SMO), Naive bayes, Decision tree, Random forest classification and Logistic Regression algorithm.The data set used in the classification algorithm obtained from the kaggle. This database contain 768 records with 9 attributes in which 8 used as input values such as Plasma glucose concentration, skin thickness, pedigree function insulin, age etc. and class as output.

Performance of the algorithm evaluated by using some performance matrix such as Accuracy, precision and recall. Performance of the data mining algorithms namely Naive bayes, Sequential Minimal optimization (SMO), Logistic Regression, Decision tree and Random forest are evaluated on the basis of the value of recall, precision, accuracy etc. Confusion matrix used to calculate the performance matrix, where P (positive) used for positive observation, N (Negative) used for negative observation, TP(True Positive) is used to correct positive prediction, FP(False Positive) is used for wrong positive prediction.TN(True Negative) used for correct negative prediction and FN used for wrong negative prediction.

5.2 Evaluation Metrics

Table 3.Confusion matrix

Confusion Matrix	A(positive predicted)	B(negative predicted)
Positive(A)	TP	FN
Negative(B)	FP	TN

Accuracy- The ratio of the sum of corrects prediction for negative and positive value and the number of total sample taken as input.

Accuracy = (TP+TN)/ (TP+TN+FP+FN)

Precision-It is define as ratio of total number of correct positive prediction and total positive prediction either it is false or true.

Precision= TP/(TP+FP)

Recall- It is define as the ratio of total number of correctly positively predicted value and the sum of correctly positively predicted value and false negatively predicted value.
Recall= TP/(TP+FN)

6. Results and Observation

The data mining algorithms: Decision tree, Naive bayes, Sequential Minimal Optimization, Random Forest and Logistic Regression are run with the pima India dataset on WEKA tool. Performance is analyzed for these algorithms for evaluating precision, accuracy and recall. Fig 6, Fig 7 and Fig 8 represent the performance of Accuracy, Recall and Precision. It is observed that decision tree, logistic regression and naive bayes have less accuracy, recall and precision , which denote the disgraceful performance while random forest and SMO are provide high accuracy, recall and precision. That means the prediction for diabetes condition is efficiently calculated by Random forest and SMO.

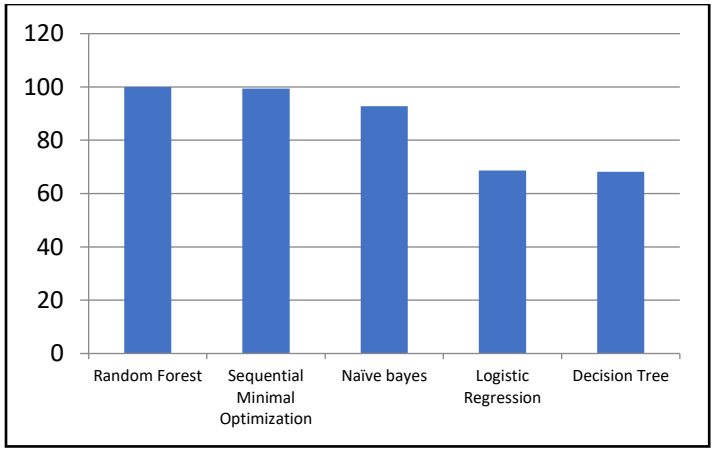


Fig6: Accuracy in percentage

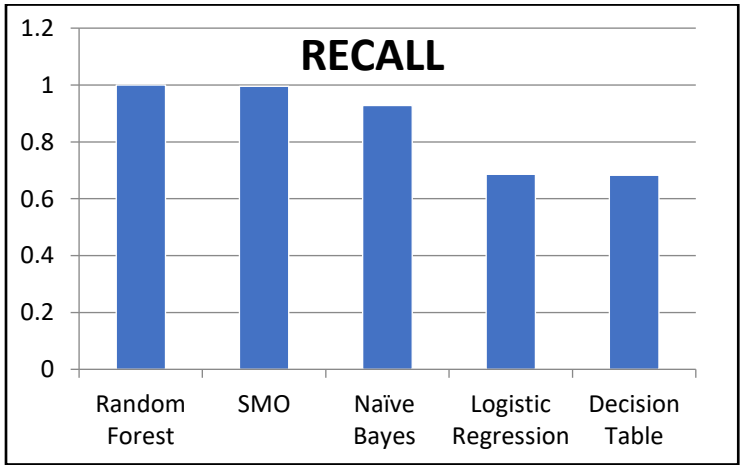


Fig7: Recall in percentage

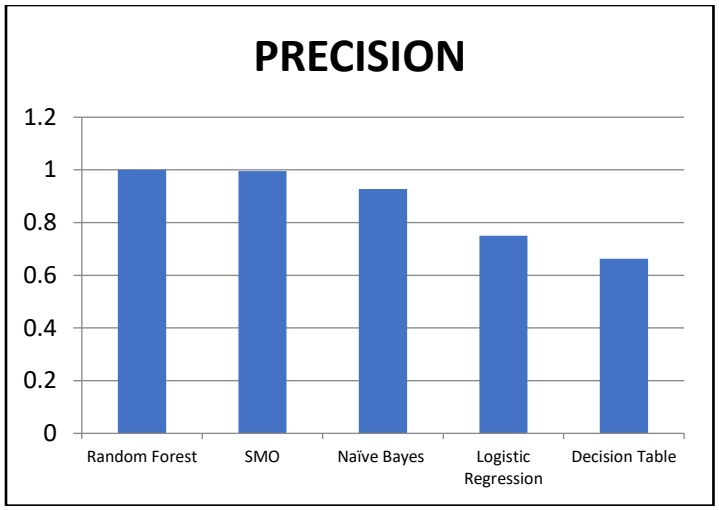


Fig 8: Precision in percentage

7.CONCLUSION

The growing rate of this chronicle disease is adding various other health issues as a risk factor. Various researches focused on that problem and conducting analytical study for the detection of health related issues with the help of machine learning algorithms. Data Analytics plays noteworthy accountability in healthcare sectors/industries. Healthcare industries collect great volume of databases. Diabetes must be prevented and cured in order to enhance the lives of all those who are impacted by it. People suffering from diabetes have more chance of getting stirred of various diseases like stroke, eye problem, heart disease, kidney disease, nerve damage, etc. Data analysis concepts are helpful in detection of complication of diabetes at the primary stage and prevent the patient form the bed effects of diabetics. This paper studies the diabetes data of the India. The dataset weighted prevalence of diabetes and pre-diabetes for ruler and urban area of the states of India. Prevalence of diabetes in percentage is just half in ruler areas as compare to urban and prevalence of pre-diabetes in rural areas is approximately 10 to 20% less than the urban areas. Glycated hemoglobin (HbA1C) value is used in the detection of diabetes and pre diabetes condition, in the human when this value is <=5.6 then its shows Non diabetes, when it is between 5.7 to 6.4 this value is denote pre- diabetes, and value >= 6.5 is denotes diabetes condition. It is observed that the diabetes mellitus affected the more population of urban areas with respect to the rural areas but the pre diabetes is condition is lightly varied in the rural and urban population as the life style and food habits of urban and rural area is quite different to each other its play a imperative role in the diabetes condition of India. Also the conversion rate of pre-diabetes to diabetes stage is higher in urban areas as compare with rural area. The experimental observation shows that random forest and SMO are surpass the logistic regression naïve base and decision tree. The accuracy of random forest is highly acceptable than others.

REFERENCES

1. Singh, A. Singh, M.N. Halgamuge, R. Lakshmikanthan, “Impact of different data types on classifier performance of random forest, naive Bayes, and K-nearest neighbor’s algorithms”, International Journal of Advanced Computer Scienceand Application, vol.8, no.12, (2017), pp.13-18.
2. Azrar A., Ali Y., Md. Awais,(2018) “Data Mining Models Comparison for Diabetes Prediction”, International Journal of Advanced Computer Science and Applications(IJACSA), 9(8), 320-323.
3. D. Dalen, G. Walker, A. Kadam “Predicting breast cancer survivability: a comparison of three data mining methods”, Artificial Intelligence in Medicine, June (2005).

4. D. Shetty, K. Rit, S. Shaikh, N. Patil, "Diabetes disease prediction using data mining", International conference on Innovation in information, embedded and communication systems (2017).
5. H. Wu, S. Yang, Z. Huang, J. He, X. Wan ,"Type 2 diabetes mellitus prediction model based on data mining ",Informatics in Medicine Unlocked, vol.10,(2017), pp.100-107.
6. J. Aravindan, "Risk Factor in patients with Type 2 diabetes in Bengaluru", World Journal of Diabetes, vol.10, no.4,(2019), pp.241-248.
7. Kavitha RK. , Singh WJ (2020). "A Study on the Effectiveness of Machine Learning Algorithms in Early Prediction of Diabetics among Patients", Biosc.Biotech.Res.Comm, 13(11), 99-104.
8. Kishore N.G., Rajesh V., Vamsi A., Reddy A, Sumedh K., Reddy TRS, (2020). "Prediction Of Diabetes Using Machine Learning Classification Algorithms", International journal of scientific Technology Research 9(1) ,1805-1808.
9. Kritika Sinha, Sunita Kushwaha, " Role of Data Mining Techniques in the Health Sectors", International Conference on Innovations in Management, Science and Technology (ICIMST) , Dept. of CS and Electronics, University of Science and Technology Meghalaya (USTM), India and American Institute of Management and Technology (AIMT), USA, 5-6 aug, 2022.
10. M.S. Kadam, I.W.Ghindawi, D.E. Mhawai,"An Accurate Diabetes Prediction System Based on K-means Clustering and Proposed Classification Approach", International Journal of Applied Engineering Research,vol.13,no.6, (2018), pp-4038-4041.
11. N.Sneha, T.Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection",Journal of big data, (2019), pp.2-19.
12. Nazim R. , Mustapha A. , Syed I., Syed Z., Abd W. & Helmy M. S, (2019). "Analyzing Diabetic Data using Classification" ,Journal of Physics: Conference Series. 1529. 022105. 10.1088/1742-6596/1529/2/022105, 1-7.
13. Pradeepa R, Mohan V(2021), " Epidemiology of type 2 diabetes in India", Indian J Ophthalmol, Volume 69 Issue 11, 2021.
14. Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju," Predicting Diabetes Mellitus with Machine Learning Techniques",Frontiers in Genetics,6 November (2018).
15. Ramachandro M. & Bhramaramba R, (2018). "Type 2 Diabetes Classification and Prediction Using Risk Score", International Journal of Pure and Applied Mathematics, volume 119, issue 15, 1099-1111. (2018).
16. S. Palaniappanand, R. Awang, "Intelligent heart disease prediction system using data mining techniques," International conference of Computer systems and applications, (2008).
17. S.A. Pattekari, A. Parveen, "Prediction system for heart disease using Naive Bayes", International Journal of Advanced Computer Math Science, vol.3, no.3, (2012), pp.290-294.
18. Seema Abhijeet Kaveeshwar, Jon Cornwall, "The current state of diabetes mellitus in India", Australas Med J , vol 7 Issue 1, 2014 .
19. Sisodia D., Sisodia D.S., (2018). "Prediction of Diabetes using Classification Algorithms", International Conference on Computational Intelligence and Data Science (ICCIDS 2018), The NorthCap University,Gurugram.
20. T.M. Ahmed," Using data mining to develop model for classifying diabetic patient control level based on historical medical records," Journal of Theoretical and applied information Technology,Vol.87, no.2, (2016), pp.316-323.
21. T.M. Allam, M. A. Iqbal, A.Yasir, A. Wahab,S. Ijaz, T.I. Baig, A. Hussian, M.A. Mallik, M.M. Raza, S. Ibrar, Z. Abbas,"A model for early prediction of diabetes", Informatics in medicine unlockedvol.16, (2019), pp.1-6.
22. V.A. Kumari, R. Chitra," Classification of diabetes disease using support vector machine", International Journal Engineering Research and Application, Vol.3, no.2, (2013), pp.1797-1801.

23. Z. Wei & Guangjian Y, (2015), "The Research on Analyzing Risk Factors of Type 2 Diabetes Mellitus Based on Improved Frequent Pattern Tree Algorithm" , Proceedings of the 2015 International Conference on Materials Engineering and Information Technology Applications.