DOI: https://doi.org/10.36094/sc.v89.2024.Prediction_of_Diabetes_Using_Machine.Mondal_Choudhury.283

ARTICLE

Sci. and Cult. 90 (7–8) : 283-284 (2024)

PREDICTION OF DIABETES USING MACHINE LEARNING MODELS

SUSHMITA MONDAL* AND J. PAUL CHOUDHURY*

Diabetes is the leading causes of blindness, stroke, kidney and heart failure. It's necessary to find the cause of disease and certain preventive measures can be taken in advance so that the concentration of disease may be minimized quickly. Machine learning models such as Random Forest, Decision tree, Naïve Bayes, LR, SVM, KNN Classifier, Gradient Boosting with hypermeters have been used. The selection of particular model has been made based on accuracy and Precision, Re-call, F1-Score, support have also been calculated.

Introduction

Diabetes is the most common and fasts growing human diseases in worldwide. Machine Learning is a revolutionizing healthcare, and diabetes prediction is prime example. Early detection allows for preventative measures, potentially reducing the burden of this chronic disease. The dataset[1] comprises of eight attributes collected from 768 patients. These attributes are pregnancies, glucose, blood-pressure, skin thickness, insulin, BMI, diabetes pedigree function, age. Based on these attributes, eight machine learning techniques have been used to predict the disease.

Literature Review

Yasodha et al[2] used classification on a diabetic dataset with 200 instances and nine attributes from hospital records, classified with Naïve Bayes, J48, REP Tree, and Random Tree. J48 showed the best accuracy at 60.2%. Aiswarya et ai[3] used Decision Tree (J48) and Naïve Bayes on the PIMA dataset achieving 78% and 79.5% accuracy respectively. Gupta et al[4] compared JRIP, J-graft, and BN algorithms, findings J-graft had the highest accuracy at 81.3% Lee et[5] applied CART Decision Tree with a resample filter, addressing class imbalance to improve model accuracy.

Result

Model wise accuracy has been furnished in table 1. Model wise precision, recall, f1 score, support have been furnished in table 2. Model wise true positive, true negative, false positive and false negative have been furnished in table 3.

Table 1: Model Wise Accuracy

S. No	Method	Accuracy
1.	Random Forest Classifier	78%
2.	Decision tree	76%
3.	Naïve Bayes	93%
4.	Logistic Regression	78%
5.	Support Vector Machine	77%
6.	KNN Classifier	70%
7.	Gradient Boosting Classifier	70%

From table-1, it has been observed that Naïve Bayes algorithm has shown excellent performance among other models. From table-2 it has been seen that the values of precision, recall, f1score are larger than other models and

* Narula Institute of Technology, Kolkata.
e-mail: sushmitam9799@gmail.com

Table-2: Model wise Precision, Recall, F1-score & Support Values

Models Name	Precision	Recall	F1-Score	Support
Random Forest Classifier	0.80	0.90	0.85	157
Decision Tree	0.80	0.86	0.83	157
Naïve Bayes	0.94	0.95	0.94	151
Logistic Regression	0.79	0.90	0.84	146
Support Vector Machine	0.72	0.89	0.80	150
KNN Classifier	0.72	0.89	0.80	150
Gradient Boosting Classifier	0.72	0.89	0.80	150

Table-3: Confusion Matrix of all Models

Models Name	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
Random Forest Classifier	120	50	30	31
Decision Tree	138	40	30	19
Naïve Bayes	143	71	9	8
Logistic Regression	129	50	30	22
Support Vector Machine	129	43	38	21
KNN Classifier	120	50	30	31
Gradient Boosting Classifier	120	50	30	31

from table-3, the values of TP and TN are more in Naïve Bayes algorithm than other models indicating superiority of Naïve Bayes algorithm than other models.

Conclusion

Naïve Bayes algorithm is simple to implement. It requires no iterations since the probabilities can be directly computed. If the conditional independence assumption holds, it can give good results. In this research work it has shown that Naïve Bayes algorithm is excellent as compared to other machine learning algorithms. ❐

References

1. https://www.dropbox.com/s/uh7o7uyeghq.
2. Aljumah, A.A., Ahamad, M.G., Siddiqui, M. K., 2013. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University – Computer and Information Sciences 25, 127 - 136. doi:10.1016 / j. jksuci.2012.10.003.
3 R. Arora, Sumon, 2012. Comparative Analysis of Classification on Different Datasets using WEKA. International Journal of Computer Applications 54, 21-25. doi:10.5120/8626-2492.
4. MP. Bamnote, G.R., 2014. Design of Classifier for Detection of Diabetes Mellitus Using Genetic programming. Advances in Intelligent Systems and Computing 1, 763-770. doi:10.1007/978-3-319-11933-5.
5. Choubey, D.K., Paul, S., Kumar, S., Kumar, S., 2017. Classification of Pima Indian diabetes dataset using Naïve Bayes with genetic algorithm as an attribute selection, in: Communication and Computer System: Proceedings of the International Conference on Communication and Computer System (ICCCS 2016), pp. 451 – 455.