ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

DIABETES DISEASES USING MACHINE LEARNING ALGORITHMS

[1]P. MOUNIKA, [2]VASA ISAIAH

[1](Assistant Professor), MSC, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MSC, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

Abstract

This paper deals with the prediction of Diabetes Disease by performing an analysis of five supervised machine learning algorithms, i.e. K-Nearest Neighbors, Naive Baye, DecisionTree Classifier, Random Forest and Support Vector Machine.Further, by incorporating all the present risk factors of the dataset, we have observed a stable accuracy after classifying and performing cross-validation. We managed to achieve a stable and highest accuracy of 76% with KNN classifier and remaining all other classifiers also give a stable accuracy of above 70%. We analyzed why specific Machine Learning classifiers do not yield stable and good accuracy by visualizing the training and testing accuracy and examining model overfitting and model underfitting. The main goal of this paper is to find the most optimal results in terms of accuracy and computational time for Diabetes disease prediction.

1. INTRODUCTION

In this day and age, one of the most notorious diseases to have taken the world by storm is Diabetes, which is a disease which causes an increase in blood glucose levels as a result of the absence or low levels of insulin. Due to the many criterion to be taken into consideration for an individual to harbour this disease, it’s detection and prediction might be tedious or sometimes inconclusive. Nevertheless, it isn’t impossible to detect it, even at an early stage. Federation- IDF). 79% of the adult population were living in the countries with the low and middle-income groups. It is estimated that by the year 2045 approx. 700 million people will have diabetes (IDF).

Diabetes is increasing day by day in the world because of environmental, genetic factors. The numbers are rising rapidly due to several factors which includes unhealthy foods, physical inactivity and many more.

Diabetes is a hormonal disorder in which the inability of the body to produce insulin causes the metabolism of sugar in the body to be abnormal, thereby, raising the blood glucose levels in the body of a particular individual. Intense hunger, thirst and frequent urination are some of the observable characteristics. Certain risk factors such as age, BMI, Glucose Levels, Blood Pressure, etc., play an important role to the contribution of the disease.In the Fig. 1 we can see that the number of cases is rising every year and there is not slowing down in the active cases. It is a very crucial thing to worry as diabetes has become one of the most dangerous and fastest diseases to take the lives of many individuals around the globe. Machine Learning is very popular these days as it is used everywhere, where a large amount of data is present, and we need some knowledge from it. Generally, we can categorise the Machine Learning algorithms in two types but not limited to-

- Unsupervised Learning: In unsupervised learning, the information is not labelled and also not trained. Here, we just put the data in action to find some patterns if possible.
- Supervised Learning: In supervised learning, we train the model based on the labels attached to the information and based on that we classify or test the new data with labels.

EXISTING SYSTEM

In [2], they have used the WEKA tool for data analytics for diabetes disease prediction on Big Data of healthcare. They used the publicly available dataset from UCI and applied different machine learning classifiers on it. The classifiers which they incorporated are Naive Bayes, Support Vector Machine, Random Forest and Simple CART.

Their approach starts with accessing the dataset, preprocess it in Weka tool and then did the 70:30 train and test split for applying different machine algorithms. They did not go with the cross-validation step as it is imperative to get the optimal and accurate results as well.

The authors in [3], also used the publicly available dataset named as Pima Indians Diabetes Database for performing their experiment. Their framework of performing the prediction starts with the dataset selection and then with data pre-processing.. As they incorporated different evaluation metrics, they did compare the different performance measure and comparatively analyzed the accuracy. The highest accuracy achieved with their experiment was 76.30%. Like [2] they have also not practised Cross-validation.

In [4], the authors proposed the neural network-based diabetes disease prediction on Indians Pima Diabetes Dataset. They have used several hidden layers to find patterns in the data, and with the help of those patterns, they predicted the outcome. They name their proposed algorithms as ADAP, which is a custom neural network with multiple partitions and with the set of association weights and units. They managed to achieve a crossover point for sensitivity, and specificity at 0.76 and are trying to precise their result in future.

Disadvantages

There are no techniques and models for analyzing large scale datasets in the existing system.

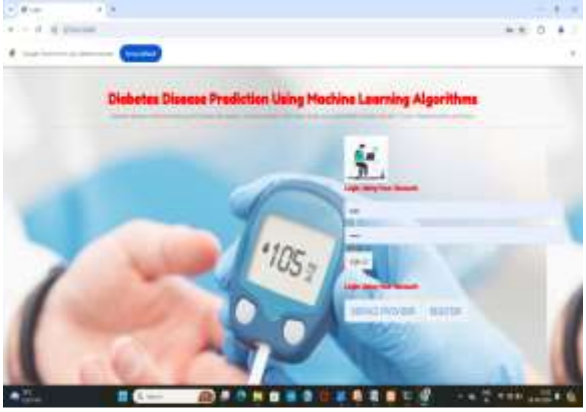
There is no facility for diabetes dataset in collaboration with a hospital or a medical institute and will try to achieve better results.

PROPOSED SYSTEM

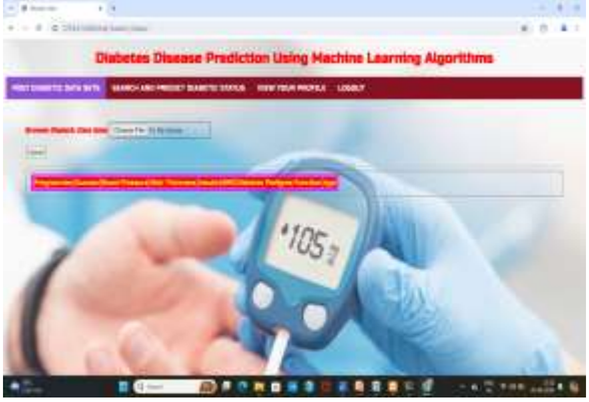
To perform our experiment, we have used a publicly available dataset named as Pima Indians Diabetes Database [4]. This dataset includes a various diagnostic measure of diabetes disease. The dataset was originally from the National Institute of Diabetes and Digestive and Kidney Diseases. All the recorded instances are of the patients whose age are above 21 years old.

Advantages

The system more effective due to fitting datasets for different ML Models by Applying Machine Learning Algorithms.


Diabetes Disease Prediction Using Machine Learning Algorithms

The Early determination of a disease can be made possible through machine learning by studying the characteristics of an individual in the proposed system


Diabetes Disease Prediction Using Machine Learning Algorithms
POST DIABETIC DATA SETS
SEARCH AND PREDICT DIABETIC STATUS
VIEW YOUR PROFILE
LOGOUT

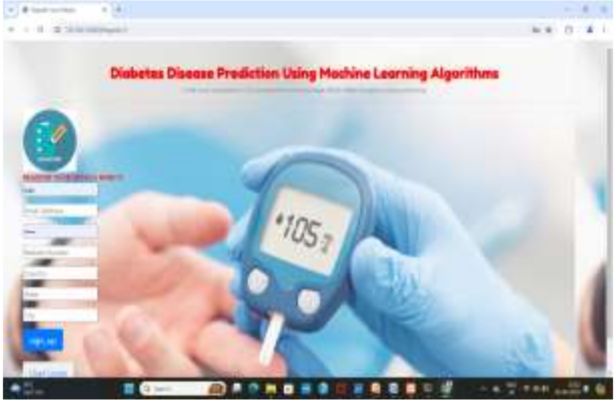
3. SCREENSHOTS

LOGIN PAGE

UPLOAD PAGE:-

SSREGISTER PAGE:-

PROFILE VIEW PAGE:-


Diabetes Disease Prediction Using Machine Learning Algorithms

4. CONCLUSION

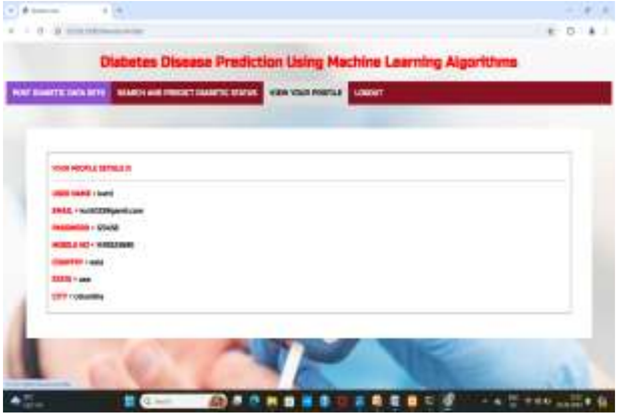
One of the significant impediments with the progression of technology and medicine is the early detection of a disease, which is in this case, diabetes. However, in this study, systematic efforts were made into designing a model which is accurate enough in determining the onset of the disease. With the experiments conducted on the Pima Indians Diabetes Database, we have readily predicted this disease. Moreover, the results achieved proved the adequacy of the system, with an accuracy of 76% using the K-Nearest Neighbours classifiers. With this being said, it is hopeful that we can implement this model into a system to predict other deadly diseases as well. There can be room for further improvement for the automation of the analysis of diabetes or any other disease in the future.

In future, we will try to create a diabetes dataset in collaborateon with a hospital or a medical institute and will try to achieve better results. We will be incorporating more Machine Learning and Deep learning models for achieving better results as well.

REFERENCES

[1] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova, J. E. Shaw, D. Bright, and R.Williams, “Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045:Results from the international diabetes federation diabetes atlas, 9th edition,” *Diabetes Research and Clinical Practice,* vol. 157, p. 107843, 2019.

[2] A. Mir and S. N. Dhage, “Diabetes disease prediction using machine learning on big data of healthcare,” in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–6.


Diabetes Disease Prediction Using Machine Learning Algorithms

YOUR PROFILE DETAILS !!

USER NAME : [illegible]

EMAIL : [illegible]

PASSWORD : [illegible]

MOBILE NO : [illegible]

COUNTRY : [illegible]

STATE : [illegible]

CITY : [illegible]

[3] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Computer Science*, vol. 132, pp. 1578 – 1585, 2018, international Conference on Computational Intelligence and Data Science. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050918308548

[4] J. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, "Using the adap learning algorithm to forcast the onset of diabetes mellitus," *Proceedings - Annual Symposium on Computer Applications in Medical Care*, vol. 10, 11 1988.

[5] P. S. Kohli and S. Arora, "Application of machine learning in disease prediction," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 2018, pp. 1–4.

[6] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, St´efan van der Walt and Jarrod Millman, Eds., 2010, pp. 56 – 61.

[7] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R'ıo, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and ´ Edouard Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, p. 28252830, 2011.