



## Machine Learning Algorithms to Predict Diabetes

<sup>1</sup>Pranav Sharma, <sup>2</sup>Prashant Singh, <sup>3</sup>Amit Kumar Pandey, <sup>4</sup>Dhyanender Jain

<sup>1,2,3,4</sup> Department of Information Technology

<sup>1,2,3,4</sup> Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India.

pranav001.shr@gmail.com<sup>1</sup>, prashant.ert@gmail.com<sup>2</sup>, amitpandey33@gmail.com<sup>3</sup>,  
dhyanendra.jain@gmail.com<sup>4</sup>

### Abstract

Diabetes occurs when the sugar level rises or the solution that controls sugar level is not enough. Insulin level go down. Weight gain can also cause it. It can also occur due to excess weight. Later, you may experience symptoms such as vision problems, headaches and other issues. When diabetes causes an increase in blood glucose levels (RBC), it can affect the ability of white blood cells to function properly. It is also possible for diabetes to spread from one patient to another (i.e. If covid sessions are not being done, the consequences of diabetes can become more severe for patients with diabetes. There are two methods to treat diabetes. We strive to help diabetes be predicted in an effective manner earlier.

**Keywords:** Machine Learning, Prediction, Support Vector Machine, Random Forest, KNN.

### Introduction

The goal of this research is to accurately predict diabetes in every case. Our lifestyle is the primary reason for diabetes patients growing. This cannot be changed but we can get treatment immediately. Diabetes is a long-term disease that can lead to an epidemic of healthcare problems worldwide. Diabetes is a disease that is caused by an increase or decrease in insulin levels. Machine learning algorithms and methods can help answer common questions. The aim of this system to predict diabetes in the early stages is. The project predicts using different methods: support vector machine; KNN; random forest. We have also set out to create a more effective method for early detection of diabetes.

### Literature Survey

Anjali and Veena Vijayan V. present the first Diabetes prediction using various machine learning approaches. Diabetes prediction using ensembling[2] different machine learning classifier. We propose a new approach to predicting diabetes using the PIMA Indian diabetic database. They propose a classification using ML models to improve the prediction of diabetes.

Diabetes prediction with machine learning algorithm[3] using feature selection and dimensionality minimization. Diabetes prediction has been published in many ways over the years. The frame work for machine learning was presented by where the authors implemented



the following: adaboost;random forest;guassian Process classification Support Vector Machine linear discriminant Analysis quadratic discriminant Analysis, artificial neural Network, naive Bayes, logistic regression and decision tree with various dimensionality reductions and cross validation.

## Prediction Model

The developed module measures blood glucose level, sugar, and based upon age, pregnancies, and other factors. PIMA Indian diabetic data is used as input. Next, we will preprocess the data by extracting features. After training the system on the feature we want for prediction, we then need to classify input into parameter such as age, sugar glucose, and hemoglobin. We then need to test the data to see if it conforms to our provided feature. These models are: support vector machine (random forest), and k-nearest neighbor. The system architecture will help in the future diagnosis of diabetes.

Based on a train dataset, it will also predict whether the person has diabetes.

Steps to Predict –

1. Data collection
2. Data defining
3. Pre-processing
4. Model-building
5. Result/analysis
6. Algorithm
7. Libraries Import
8. Importing dataset
9. Training Dataset
10. Test Dataset
11. Algorithms
12. Evaluation
13. Comparison of results

### **K-nearest neighbors -**

KNN is a simple but supervised machine learning algorithm. KNN algorithm has higher accuracy and is used for classification and regression. The model in that data is the reference for the future, so it does not require a training period.

### **Support Vector Machine**

Support vector machine is an algorithm that's used to compare data points from a large number of features. There are two types: linear and nonlinear support vector machines (Svm). The support vector machine can be used to cluster. Many ways are now possible to consider unsupervised learning.

### **Random Forest -**

This algorithm can be used for both regression and classification. It contains decision tree classifier. This algorithm resolves the problem of fitting training set to prefer over a decision

tree. It is mostly used for classification. This classifier provides final prediction by collecting the majority of the data.

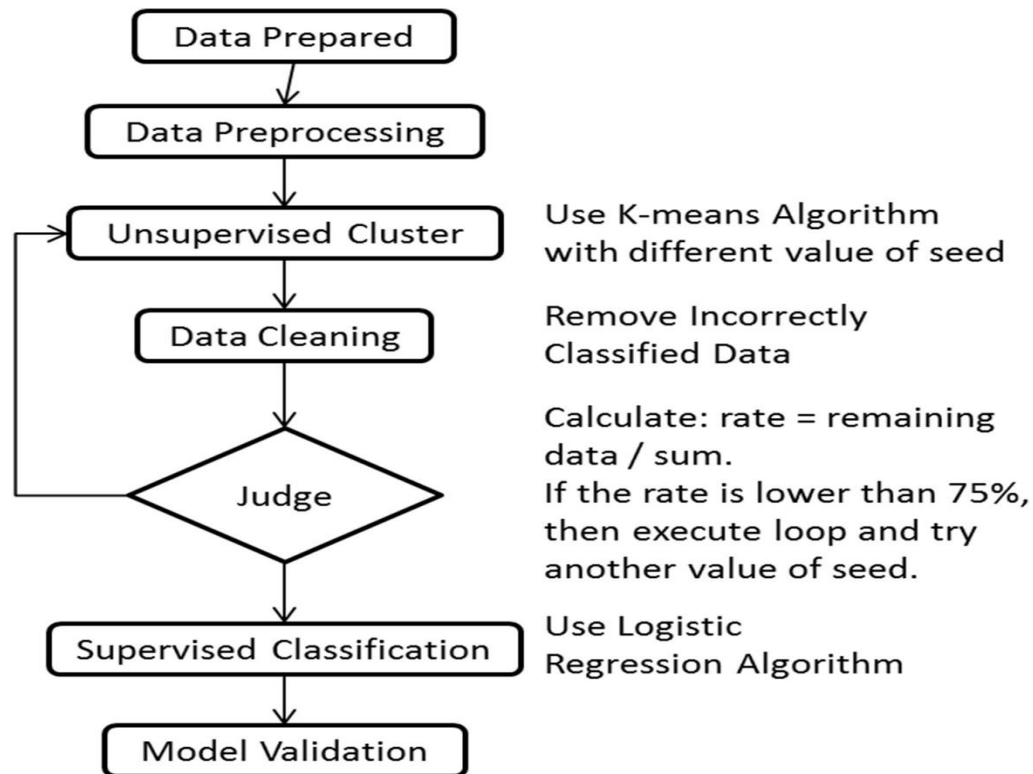


Figure 1: Algorithm Model (1)

### Conclusion

The earlier stages of diabetes prediction are very important since it is not possible to detect the disease at its onset. The system module, which is used parameters such as sugar level and glucose, age and many other factors to predict diabetes, is only going to be used to determine whether the patient/user will develop diabetes. This model assists doctors/physicians to initiate treatment as soon possible. Three different machine learning techniques were used for prediction (i.e. Support vector machine (or random forest), KNN, or random forest. The random forest method is more accurate and close to the prediction. Diabetes is a major health concern in modern society. This paper summarizes a prediction technique for this affliction.

### References

1. Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang, "Type 2 diabetes mellitus prediction model based on data mining", Informatics in Medicine Unlocked, Volume 10, 2018, Pages 100-107, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2017.12.006>.



2. Rajesh, K., and V. Sangeetha. "Application of data mining methods and techniques for diabetes diagnosis." *International Journal of Engineering and Innovative Technology (IJEIT)* 2, no. 3 (2012).
3. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, et al., "Prediction of incident diabetes mellitus in middle-aged adults: the Framingham offspring study". *Arch Intern Med.* 2007, 167:1068–74.
4. Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *Kaohsiung J Med Sci.* 2013, 29(2):93–9.
5. Dost Muhammad Khan<sup>1</sup>, Nawaz Mohamudally, "An Integration of K-means and Decision Tree (ID3) towards a more Efficient Data Mining Algorithm" *Journal Of Computing*, 3 (12) (December 2011)
6. Humar Kahramanli, Novruz Allahverdi, "Design of a Hybrid System for the Diabetes and Heart Disease" *Expert Systems with Applications: An International Journal*, 35 (1-2) (July, 2008)