Diabetes Prediction using Machine learning Techniques

Priya Gandhi, Dr Gayatri S Pandi

Priya Gandhi, Post Graduate Scholar, Post Graduate Department LJ University, Ahmedabad

Dr Gayatri S Pandi, Head of Department, Post Graduate Department, LJ University

ABSTRACT:

Diabetes is a pretty rare condition in today's world, but it is becoming more common by the day. Diabetic is a disease caused by a high glucose level in our bodies. Diabetes should not be disregarded, and treatment should begin as soon as possible. Diabetes can lead to serious complications such as heart disease, kidney disease, and damage to other body parts. The major goal of this paper is to improve the accuracy and precision of diabetic prediction. Machine learning algorithms such as naive bayes classifier, SVM, decision tree, and KNN are used. PIMA Indian dataset is taken into account for experimental analysis. According to the findings, MLP classifies accuracy between 80 and 86 percent, whereas LSTM improves accuracy to 87.26 percent for diabetes. The goal of this research is to improve the accuracy of diabetes prediction by enhancing precision and accuracy.

Keywords: Classification, Navie Bayes Classifier, Support Vector Machine, Decision Tree, KNN.

Introduction:

Diabetes Mellitus is a metabolic condition that results in excessive blood sugar levels. Insulin is a hormone that transports sugar from blood cells to be stored or used for energy. Diabetes is caused by the body's inability to produce enough insulin. Diabetes-related elevated blood sugar can harm your nerves, eyes, kidneys, and other organs if left untreated.
.
There are a two different types of diabetes-

Type 1 Diabetes:- Type 1 diabetes is a disorder in which your immune system attacks and destroys your pancreas' insulin-producing cells. They're also known as beta cells. Juvenile diabetes refers to the fact that it is most commonly diagnosed in children and teenagers. In this case, your body does not react to insulin as it should. Extreme thirst, increased hunger, dry mouth, frequent urination, and unexplained weight loss are the most common symptoms.

Type 2 Diabetes:- Type 2 diabetes is a chronic illness that prevents your body from properly utilising insulin. Insulin resistance is a term used to describe the condition of people with type 2 diabetes. People in their middle or later years are more likely to get this kind of diabetes. Approximately 84 million people have prediabetes, which means their blood sugar is high but not high enough to be diagnosed with diabetes. The most common symptoms are extreme thirst, exhaustion, weight loss without attempting, and an increase in infection. When you have type 2 diabetes, your pancreas produces the hormone insulin.

Related Works:

The goal of this project is to create a system that can combine the results of various machine learning techniques such as Navie Bayes Classifier, Decision Tree, SVM, and KNN to conduct early diabetes prediction for a patient with a greater accuracy.

Machine learning is the study of computer algorithms that can learn and develop on their own through experience and data. Artificial intelligence includes it. Machine learning algorithms create a model based on training and testing data samples. To make a forecast or a choice. Machine learning is mostly utilized in computer email, speech recognition, and computer vision, among other applications. Without being expressly coded, Machine Learning programs may complete tasks. Machine Learning has two goals: one is to classify data based on models that have been established, and the other is to generate predictions about future outcomes based on these models.

Uncontrolled diabetes causes hyperglycemia, or high blood sugar, which causes catastrophic damage to many of the body's systems, including the neurons and blood vessels, over time. Diabetes-related premature mortality rates increased by 5% between 2000 and 2016. Diabetes declined in high-income countries, but it climbed during 2010-2016. Diabetes cases grew in low-income nations over both eras.

There's also a type of diabetes known as gestational diabetes. It's a type of hyperglycemia that develops when blood glucose levels rise over normal, and it's most common during pregnancy. Complications during pregnancy and delivery are more likely in women with gestational diabetes. These mothers, as well as their children, are at a higher risk of developing type 2 diabetes in the future. Prenatal screening, rather than reported symptoms, is used to identify gestational diabetes.

We hope to make better predictions with the help of this study. The goal is to achieve accuracy and precision. We're making it more accurate and precise with the help of machine learning algorithms. We want to outperform the competition in terms of accuracy. When compared to the other algorithms, the SVM and KNN algorithms had the best accuracy rate.

History Of Diabetes Hesy-Ra, an Egyptian physician, noted frequent urination as a symptom of a strange ailment that also produced emaciation in 1552 B.C., which was the first known mention of diabetic symptoms. Around the same time, ancient healers noticed that ants were drawn to the urine of people suffering from the sickness. The Greek physician Arateus characterized diabetes as "the melting away of flesh and limbs into urine" about 150 AD. Physicians began to obtain a better understanding of diabetes after then. People known as "water tasters" diagnosed diabetes by tasting the urine of people suspected of having it centuries later. Diabetes was diagnosed if the urine tasted sweet. In 1675, the word "mellitus," which means honey, was added to the name "diabetes," which means siphon, to recognize this trait. Scientists didn't develop chemical assays to detect the presence of sugar in urine until the 1800s.
.

Methods of Machine learning techniques

Supervised Learning:- This technique involves feeding a large collection of training data to an algorithm and then letting the model study the results. It also tweaks the parameters till the desired outcome is achieved. You may next put the data to the test by asking the machine to produce a prediction for a validation data set, or new data that hasn't been seen before. Prediction, regression, and classification are common applications of supervised learning.

Unsupervised Learning:- Unsupervised Learning allows a machine to examine a set of data without being supervised. Following the initial procedure, the machine seeks out a hidden pattern that connects several variables. This kind of learning can aid in the grouping of data based on statistical features. Unsupervised learning is more faster because it does not require training on huge data sets. In comparison to supervised learning, it is also easier to implement.

Regression:- For supervised machine learning, regression approaches are utilized. The purpose of regression techniques is to use a previous data set to explain or forecast a certain numeric result. The simplest and most fundamental strategy is linear regression.

Classification:- A classification algorithm can explain or forecast the value of a class. For many AI applications, classification is a necessary component. It is primarily utilized in ecommerce applications. Classification algorithms aren't restricted to just two classes; they can be applied to a large number of datasets or categories. The simplest and most fundamental classification approach is logistic regression. To estimate the probability of an event occurring, more than one input is required, as well as the usage of data. The result is a number that is most likely between one and zero.

Clustering:- Clustering algorithms are a type of unsupervised learning method. K-means, mean-shift, and expectation-maximization are three typical clustering techniques. They categorize data points based on shared or comparable characteristics. The method of grouping or clustering is extremely beneficial in business applications. When enormous amounts of data must be segmented or classified. Clustering can also be used to find patterns in large data sets that aren't visible to the naked eye.

Decision Tree:- The decision tree algorithm classifies objects by answering questions about their qualities that are located at nodal nodes. Depending on the answer, one of the branches is chosen, and a new question is presented at the next junction. Until the algorithm gets the ultimate response, which is a tree leaf. Knowledge management platform for customer service, predictive pricing, and product planning are all part of the decision tree application.

Neural Network:- Neural networks are designed to mirror the structure of the brain, with each artificial neuron connecting to many other neurons, resulting in a complex cognitive structure made up of millions of neurons. Neurons are multilayered, with one layer transmitting info to many neurons on the next layer, and so on. Finally, the data reaches the output layer, where the network creates a decision tree to identify an object and solve a problem. Deep learning is a topic of research that is named after the multilayer structure of neural networks. Most commonly utilized in corporate applications, medical image analysis, diagnostic techniques, machine translation, and fraud detection.

.**Support Vector Machine:-** SVM stands for Support Vector Machine, which is a type of supervised machine learning technique. It can be used for both classification and regression. The goal of this approach is to find a hyperplane in an N-dimensional space that categorizes data points clearly. The hyperplane's dimensions are determined by the number of features. The purpose of SVM is to find the best line or boundary to remove from the n-dimensional space into the classes so that we can classify the remaining data or new data in the future. Hyperplane is the optimal decision boundary
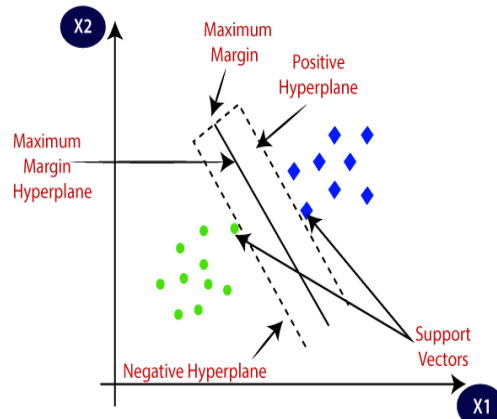
Fig:-1 Different Categories that classifies Decision Boundary or Hyperplane Of SVM

Types of SVM:- Linear And Non Linear SVM

Linear SVM:- A linear SVM classifier is used in a straight line separable dataset, which is a dataset that can be classified into two classes using a single straight line. This type of data is known as nearly straight line separable data, and the classifier used in this process is known as linear SVM classifier.

Non-Linear SVM:- Non-Linear SVM is used for data that does not develop from one level to the next level separated data, that is, data that is not classified using a straight line. This type of data is referred to as non-linear data, and the classifier used for this process is referred to as non-linear SVM.

.**Study According To The Algorithm**

According to the SVM, the best separation occurs when the hyperplane is maximized by maximizing the distance between the two boundaries. The naive bayes classifier is based on conditional probability and specifies the status of a given feature in a class without affecting the status of other features. The fundamental goal of a decision tree classifier is to predict the target class using a decision rule from prediction. Cross validation, accuracy, re-call, precision, and other factors are used in naive bayes, SVM, and decision trees
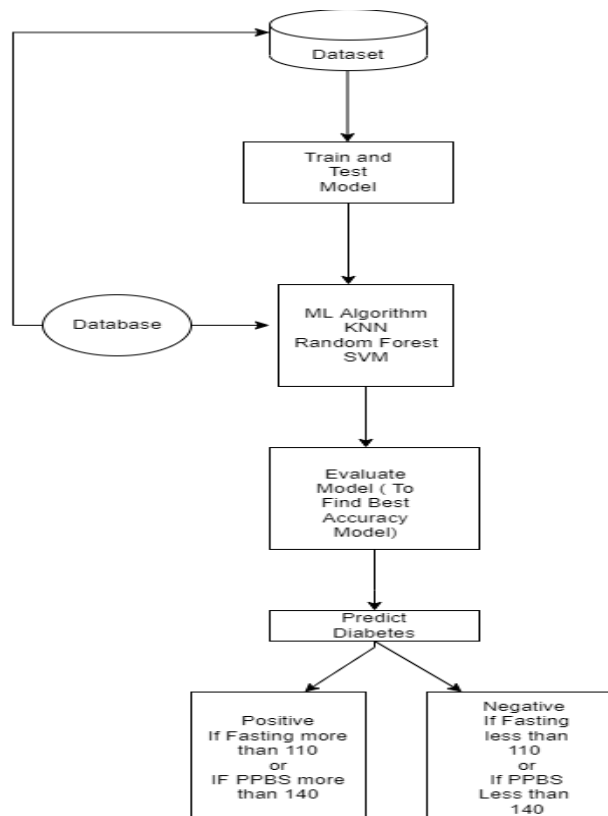
Proposed System

Fig:-2 Proposed System For Diabetic System

The suggested diabetes system is depicted in the illustration.The dataset is initially taken into consideration and is collected from the PIMA dataset. The model is trained and tested using the dataset. After that, the model is trained and tested according to the techniques used, which include KNN, Random Forest, and SVM. The method is used, and the model's best accuracy is taken into account. The SVM and KNN models are said to have the best accuracy and precision. The outcome is based on the most precise and accurate model available. After then, diabetes is a foregone conclusion. According to the results of the PPBS and fasting tests, It is positive if fasting is greater than 110 or PBBS is greater than 140; otherwise, it is negative if fasting is less than 110 or PBBS is less than 140.

Table-1 Attributes of the detailed dataset for Diabetes

| Attribute	Description
num_preg	Number of pregnancies (Numeric).
glucose_conc	Plasma glucose concentration (Numeric).
diastolic_bp	Diastolic blood pressure (mm Hg) (Numeric).
thickness	Triceps skin fold thickness (mm) (Numeric).
insulin	2-Hour serum insulin (mu U/ml) (Numeric).
Bmi	Body mass index (Numeric).
diab_pred	Diabetes pedigree function (Numeric)
age	Age (Numeric).
Diabetes	Diabetes or not diabetes (True/False).

.
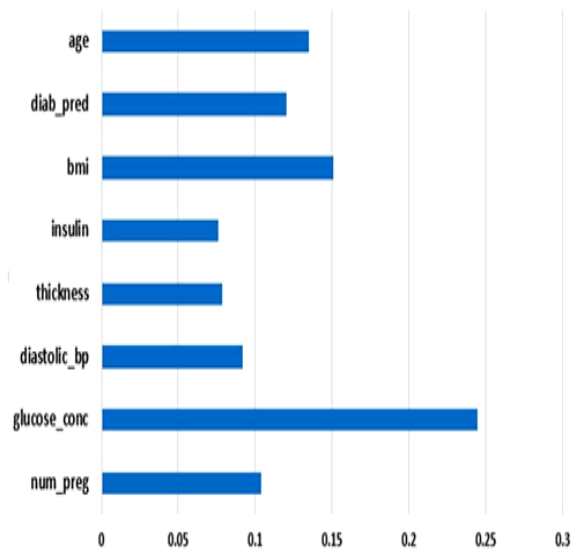
Fig:-3 Feature Importance According to the Attributes Of Diabetes

Table -2 Percentage Accuracy Results of Classification Techniques

Support Vector Machine (SVM)	K-Nearest Neighbor (KNN)
Testing set:90.23	Testing set: 79.22
Training set: 81.78	Training set: 71.98

Result and Analysis:-

The purpose of this review paper is to determine whether or not a patient will develop diabetes disease. Support vector machine and k-nearest neighbor were used in this study to study supervised machine learning classification approaches. The WEKA tool was used to conduct a number of experiments involving various classifier algorithms. Data is pre-processed using supervised classification techniques like SVM and KNN. To get the accuracy score, K-nearest neighbor is used. For training and test data sets, the accuracy score results of several classification techniques were reported in python programming. Table no. 2 shows the percentage accuracy scores.

Graph of the Result: Comparison Between Decision Tree, KNN and SVM

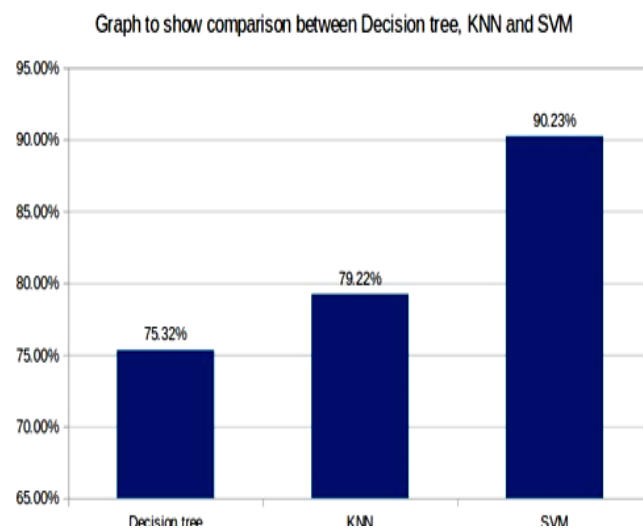Graph to show comparison between Decision tree, KNN and SVM

Fig:-4 Graph shows result of comparison between decision tree, KNN, SVM algorithm

Conclusion:

Diabetic identification at an early stage is one of the most pressing medical issues. This study or research makes a concerted attempt to help the user recognize diseases such as diabetes. This system employs machine learning techniques and algorithms such as SVM and Naive Bayes classifier. The results of the experiments are based on algorithms and accuracy, precision, and other factors.

Using machine learning techniques, the proposed system will undoubtedly aid in the prediction of diabetes. This will aid in improving its accuracy and decreasing the amount of strange data. With advanced machine learning techniques, the existing system focuses on huge datasets. The proposed system, on the other hand, is working on a number of techniques that are more accurate and are enhancing their accuracy.

References:

[1] Diabetes Prediction using Machine Learning TechniquesMitushi Soni Dept of Computer Science and Engineering Shri G.S. Institute of Technology and Science Indore, India Dr. Sunita Varma. Dept of Information Technology Shri G.S. Institute of Technology and Science Indore, India.

[2] Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications Umair Muneer Butt , Sukumar Letchmunan ,Mubashir Ali ,Fadratul Hafinaz Hassan , Anees Baqir and Hafiz Husnain Raza Sherazi School of Computer Sciences, University Sains Malaysia, Penang, Malaysia Department of Management,Information and Production Engineering, University of Bergamo, Bergamo, Italy Department of Environmental Sciences ,Informatics,and Statistics, Ca'Foscari University of Venice, Venice, Italy, School of Computing and Engineering, University of West London, London, UK

[3]https://www.healthline.com/health/diabetes 22/01/2022

[4] https://www.researchgate.com/health/diabetes 22/01/2022

 [5] Type 1 Diabetes Medically Reviewed by Poonam Sachdev on October 08, 2021

[6] Type 2 Diabetes Medically Reviewed by Michael Dansinger, MD on December 06, 2020

[7] An Efficient Prediction System for Diabetes Disease Based on Deep Neural Network Tawfik Beghriche , Mohamed Djerioui , Youcef Brik , Bilal Attallah , and Samir Brahim Belhaouari

[8] https://www.who.int/news-room/fact-sheets/detail/diabetes

[9] The History of Diabetes By Krisha McCoy Medically Reviewed by Pat F. Bass III, MD, MPHReviewed: November 3, 2009

[10] Prediction Of Diabetes Using ML Algorithms In Healthcare Author:- Muhammad Azeem Sarwar, Nasir Kamal, Wajeeha Hamid,Munam Ali Shah Publication:- IEEE Journal. Table no 1 and Fig:-3.

[11]Hassan, A.S., Malaserene, I. and Leema, A.A., 2020. Diabetes Mellitus Prediction using Classification Techniques. Int. J. Innovation. Technol. Explore. Eng, 9(5), pp.2080-2084.a

[12]Singh, D.A.A.G., Leavline, E.J. and Baig, B.S., 2017. Diabetes prediction using medical data. Journal of Computational Intelligence in Bioinformatics, 10(1), pp.1-8.

[13]Sonar, P. and JayaMalini, K., 2019, March. Diabetes prediction using different machine learning approaches. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 367-371). IEEE. Table no:-2

[14]Sisodia, D. and Sisodia, D.S., 2018. Prediction of diabetes using classification algorithms. Procedia computer science, 132, pp.1578-1585.

[15]Sarwar, M.A., Kamal, N., Hamid, W. and Shah, M.A., 2018, September. Prediction of diabetes using machine learning algorithms in healthcare. In 2018 24th International Conference on Automation and Computing (ICAC) (pp. 1-6). IEEE.

[16] https://www.geeksforgeeks.org/support-vector-machine-algorithm/ 21/01/22

[17] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm fig:-1 22/01/22

[18]DiabetesMellitus   predictionUsingClassificationTechniqueAuthor:-AbdulhakimSalun Hassan, I.Malaserene, A.Anna LeemaPublication:- IJITEE 2020 fig:-4