**Predictive Analytics on Diabetes Data using Machine Learning Techniques**

**Alternative Title:**

Prediction and detection of diabetes using web application

**Aim:**

To help doctors and practitioners in early prediction of diabetes using machine learning techniques**.**

**Abstract:**

Diabetes caused due to increase in amount of sugar or glucose which is condensed into the blood Identifying process of diabetes is the glucose and sugar levels needs to be checked before and after meal, there are fluctuations before and after meal, this whole process of patient visiting a doctor is tiresome. But in Machine Learning algorithms helps us to solve this issue. The motive of this study and research is to make use of features and to predict the likelihood of the disease, Decision Tree, Random Forest, K Nearest Neighbours, Naive Bayes and Support Vector Machine are the algorithm that have been applied to detect and predict diabetes at an early stage. A dataset of a patient’s medical record is obtained and different machine learning algorithms are applied on the dataset. Performance and accuracy of the applied algorithms is discussed and compared. Comparison of the different machine learning techniques used in this study reveals which algorithm is best suited for prediction of diabetes.

**Existing System:**

It is apparent from the truth that the occurrence of diabetes mellitus is high and the complication in the prevention of diabetes also increases. Thus, there are many patients who need the required knowledge and skills to enrich their health. In such cases, the patients are needed to visit the diagnostic center for their treatment. Because of this, they lost their time and expenses. The data of diabetic patients collected from the UCI laboratory is used to discover patterns with KNN and Support Vector Machine (SVM). The results are compared for performance and accuracy with these algorithms. It gives some moderate level of accuracy .So we use more ML algorithms to predict diabetic accurately using web framework.

**Proposed System:**

In this paper, we are using machine learning algorithms to predict diabetes disease. The client on their first login has to register themselves on the Web Application. The web Application created by Django. Once the user logins into the system he gets all the access for predicting the diabetic and by using the input such as Pregnancies, Glucose, Blood Pressure, Skin Thickness, BMI, insulin level and age based on their own. After submitting the inputs, it’s move on to the trained model for comparison. Already trained model were trained by machine learning algorithms. Algorithms used for training a dataset are K Nearest Neighbours (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF). Comparison of the different machine learning techniques used in this study reveals which algorithm is best suited for prediction of diabetes.

**Module Description:**

* Data Pre-Processing
* Machine Learning Implementation
* Classification Modeling
* Prediction

**Dataset Description:**

The objective of the dataset is to predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

**Data Pre-Processing:**

Diabetic disease data is pre-processed after collection of various records. The dataset contains much number of patient records, where some records are with some missing values. Those missing records have been removed from the dataset and the remaining patient records are used in pre-processing.

**Feature Selection**

From among the 8 attributes of the data set, one attributes pertaining to age is used to identify the personal information of the patient. The remaining 7 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of diabetes disease.

**Classification Modeling**

The Classification Algorithms to produce the best results. We are using SVM, KNN, Naïve Bayas, Logistic Regression, Decision Tree and Random Forest Algorithm to predict the diabetics using ML. The clustering of data sets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error. The K-Nearest neighbours give the high accuracy level of prediction

* Decision Trees Classifier
* Support Vector Classifier
* Random Forest Classifier
* K Nearest neighbors
* Naive Bayes

**Performance Measures:**

Preprocessed data are trained by ML algorithm and input given by the user from web application, goes to the trained dataset. After prediction the predict value Shown as an output on web application.

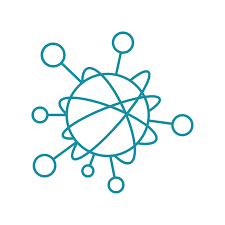
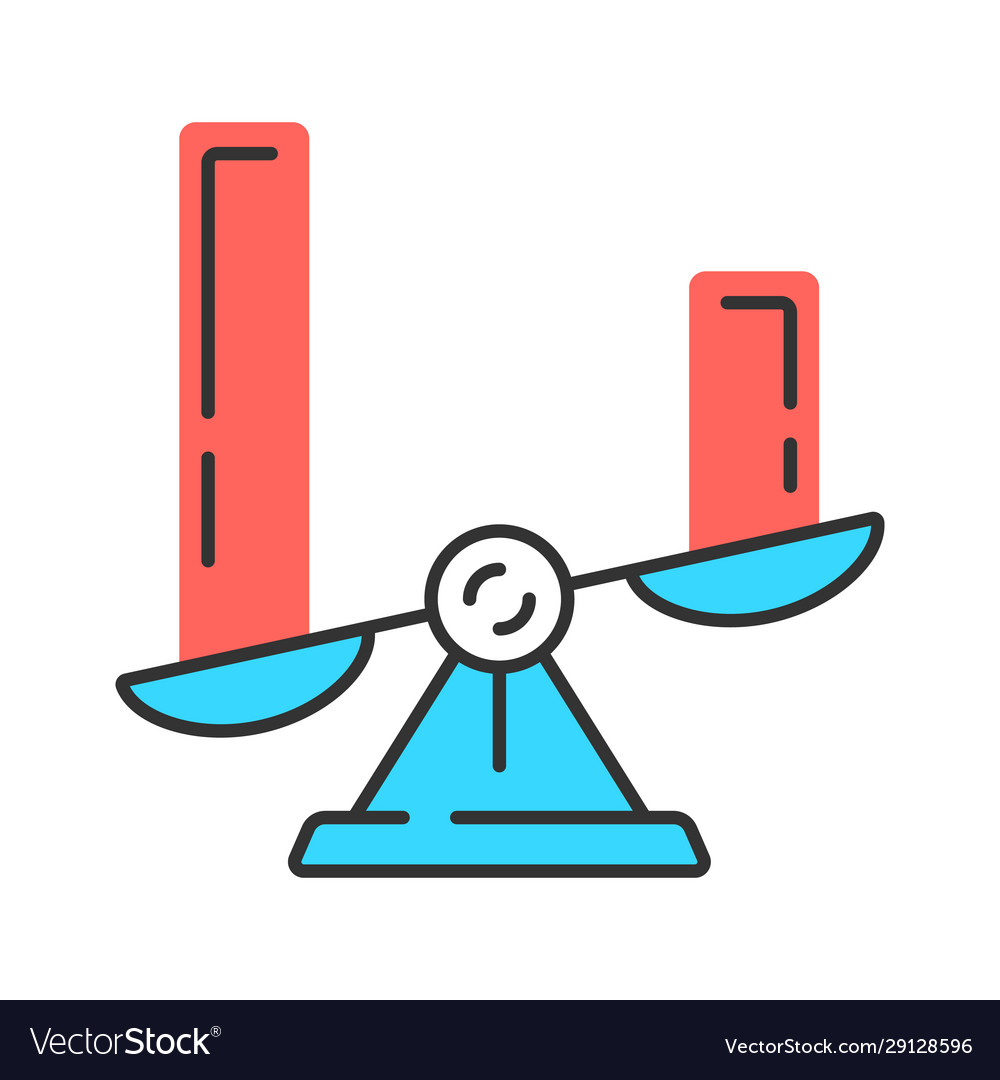
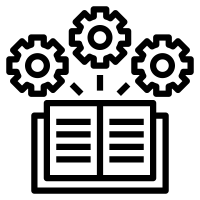
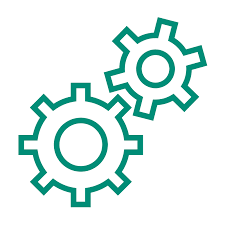
**Software Requirements:**

* Operating System : Windows 7 , 8, 10 (64 bit)
* Software : Python
* Tools : Anaconda (Jupyter Note Book IDE)

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Architecture Diagram:**



Dataset

Preprocessing

Training model

Compare to trained model

Web Application

Diabetic Prediction