**PREDICTING DIABETES IN HEALTHY POPULATON THROUGH ML**

**OBJECTIVE:**

The Main objective of this project is detected whether a patient have Diabetes or not and to know this, we used classification techniques of Decision Trees, Random Forest, Logistic Regression, ANN, SVM and hybrid model

**ABSTRACT:**

Diabetes is a disease that has no permanent cure; hence early detection is required. Data mining, machine learning (ML) algorithms, and Neural Network (NN) methods are used in diabetes prediction in our research. We used the Pima Indian Diabetes (PID) dataset for our research, collected from the UCI Machine Learning Repository. The dataset contains information about 768 patients and their corresponding nine unique attributes. We used seven ML algorithms on the dataset to predict diabetes. We found that the model with Logistic Regression (LR) and Support Vector Machine (SVM) works well on diabetes prediction. We built the NN model with a different hidden layer with various epochs and observed the NN with two hidden layers

**Keywords:** Decision Trees, Random Forest Logistic Regression, ANN, Hybrid Model

**EXISTING SYSTEM:**

The increasing growth of machine learning, computer techniques divided into traditional methods and machine learning methods. This section describes the related works of Diabetes Prediction Using Different Machine Learning Classifiers and how machine learning methods are better than traditional methods. The existing method in this project have a certain flow is used for model development. Decision Trees, Random Forest, ANN, SVC are used algorithms in existing system. But it requires large memory and result is not accurate.

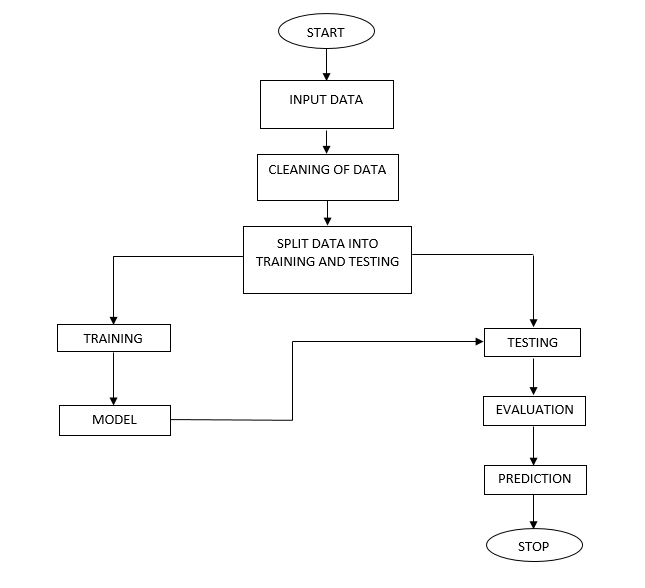
**Disadvantages:**

* Accuracy low
* Operating cost is high
* difficult to handle

**PROPOSED SYSTEM:**

Proposed several machine learning models model to classify Diabetes or not, but none have adequately addressed this misdiagnosis problem. That can be used for this purpose are Stevens Multi a Diabetes Prediction Using Different Machine Learning Classifiers. Also, similar studies that have proposed models for evaluation of such tumors mostly do not consider the heterogeneity and the size of the data Therefore, we propose a machine learning-based approach which combines a new technique of pre-processing the data for features transformation, classifications hybrid algorithm, Logistic Regression. These both ML algorithms give the best accuracy techniques to eliminate the bias and the deviation of instability and performing classifier tests based

**Flow Chart:**

****

**Fig:** Block Diagram

**Advantages:**

* Highest accuracy
* Reduces time complexity.

.

**SOFTWARE AND HARDWARE REQUIREMENTS:**

**Hardware:**

Operating system : Windows 7 or 7+

RAM : 8 GB

Hard disc or SSD : More than 500 GB

Processor : Intel 3rd generation or high or Ryzen with 8 GB Ram

**Software:**

Software’s : Python 3.6 or high version

IDE : PyCharm.

Framework : Django or flask

**LEARNING OUTCOMES:**

* About Classification in machine learning.
* About preprocessing techniques.
* About DecisionTreeClassifier.
* About Logistic Regression.
* About Random Forest.
* About SVM
* About ANN
* About Hybrid model
* Knowledge on PyCharm Editor.