**CLASSIFICATION AND PREDICTION OF DIABETES RISK USING MACHINE LEARNING TECHNIQUES WITH DATA MINING ALGORITHMS.**

**Abstract:**

Diabetes diseases are the one cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of Diabetes diseases and continuous supervision of clinicians can reduce the mortality rate. However, accurate detection of Diabetes in all cases and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. In this study, a tentative design of a machine learning based Diabetes disease prediction system had been proposed to detect impending Diabetes disease using Machine learning techniques. For the accurate detection of the Diabetes disease, an efficient machine learning technique should be used which had been derived from a distinctive analysis among several machine learning algorithms. Machine Learning is used across many spheres around the world. The healthcare industry is no exception. Machine Learning can play an essential role in predicting presence/absence of Locomotors disorders, Diabetes more. Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment per patient basis. And in addition we also predicting the diseases the machine learning. Algorithms like Random Forest, Support Vector Machine, Naïve Bayes and etc., are used to build a machine learning model. Here discusses and compares the various data analytics techniques available for the Diabetes prediction. However, the selection of the appropriate algorithm from the pool of available algorithms imposes challenge to the researchers with respect to the chosen Diabetes. The accuracy of training model should be higher and error rate should be minimum.

**Literature Survey 1**

# Title: Predicting Diabetes Disease at Early Stages using Machine Learning: A Survey

**Year:** 2020

**Author:** Rahul Katarya, and Srinivas Polipireddy

**Abstract:** Predicting and detection of Diabetes disease has always been a critical and challenging task for healthcare practitioners. Hospitals and other clinics are offering expensive therapies and operations to treat Diabetes. So, predicting Diabetes disease at the early stages will be useful to the people around the world so that they will take necessary actions before getting severe. Diabetes disease is a significant problem in recent times; the main reason for this disease is the intake of alcohol, tobacco, and lack of physical exercise. Over the years, machine learning shows effective results in making decisions and predictions from the broad set of data produced by the health care industry. Some of the supervised machine learning techniques used in this prediction of Diabetes disease are artificial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes) (NB) and k-nearest neighbor algorithm. Furthermore, the performances of these algorithms are summarized.

**Literature Survey 2**

# Title: Prediction of Diabetes Disease Using Machine Learning

**Year:** 2018

**Author:** Aditi Gavhane

**Abstract:** With the rampant increase in the Diabetes stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a Diabetes stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a Diabetes disease. Thus we propose to develop an application which can predict the vulnerability of a Diabetes disease given basic symptoms like age, sex, pulse rate etc. The machine learning algorithm neural networks has proven to be the most accurate and reliable algorithm and hence used in the proposed system.

**Existing system:**

In existing system, the medical teams diagnose the disease from the patient’s medical history, symptoms, a physical exam and test. It’s not possible to predict Diabetes by simply seeing it, but preventing a Diabetes by taking physical test it is possible. This in this existing system they use only machine techniques to predict whether the patient or person having Diabetes or not and in this study they implement the model with single data mining without web application.

**Disadvantage**:

1.uses single trained model

2.only able to run in python platform.

3.less accuracy with machine learning model.

**Proposed system:**

In this proposed system, we adopt the artificial intelligent concept known as machine learning in order to predict whether the patient or person is affected by Diabetes or not with the use result thus in this proposed we are predicting Diabetes disease and also .Here, machine learning technique supervised learning is used for predication of outcome for both Diabetes . Based on various inputs like smoking, Diabetes rate, BP , etc…. in order to predict the Diabetes . Here there is no need of doctor after the physical test because our machine learning model itself will predict the outcome and it is the one of the main advantage carried by this study and also in this study we develop the machine learning model and embed with web application so that it can be use it in the real time application.

**Advantages**:

1. Increasingly Accurate Dataset and Model.

2. Separate Model for each stage. Initial stage and final stage

3. Interactive Model and Interfaced with Web Application.

4. Recommending Hospitals and Doctors

5. Able to run in mobile.

**SYSTEM REQUIREMENTS :**

**Functional requirement**

Language : Python

Packages  : Pandas, Numpy , Sklearn ,Flask

Framework : Anaconda

Front End : HTML

Back End : Machine Learning Mode

**Hardware Requirements :**

Operating System : Windows OS

Processor : i3 or higher

Ram : 4 GB or higher

IDE : Anaconda

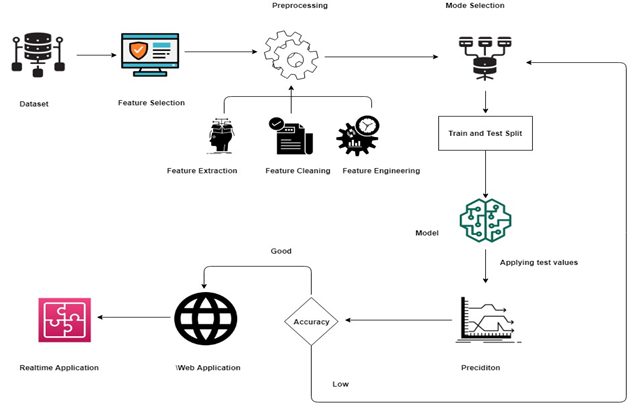
**MODULES:**

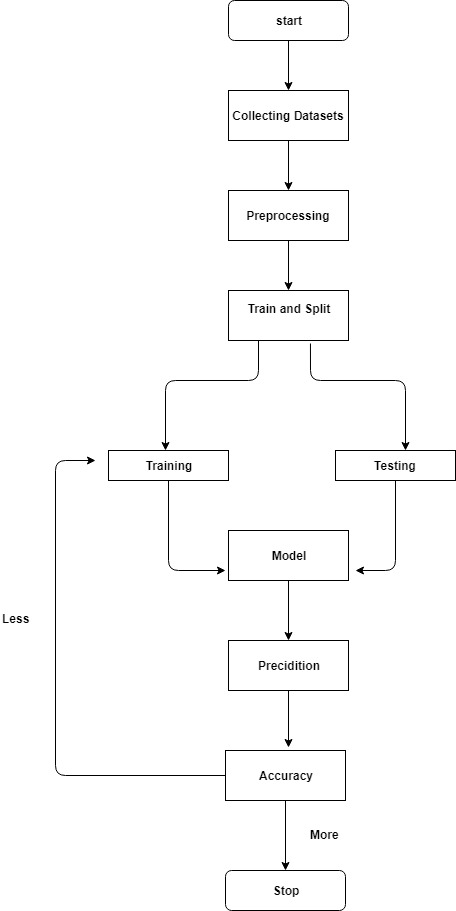
1. Data Collection and Preprocessing.
2. Creating model.
3. Creating heart diseases module.
4. Interface with UI.

**WEB BASED MODULES:**

1. Login and registration
2. Module
3. Heart disease module
4. Hospital and doctor suggestion

**SYSTEM ARCHITECTURE:**



**DATAFLOW DIAGRAM:** 

**SOFTWARE REQUIREMENT**

Python

Numpy

Pandas

Data rrame

SKLearn

Matplotlib

**CONCLUSION**

due to Diabetes, it has become mandatory to develop a system to predict Diabetes effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of Diabetes. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 90.16% for prediction of Diabetes disease. In future the work can be enhanced by developing a web application based on the Random Forest algorithm as well as using a larger dataset as compared to the one used in this analysis which will help to provide better results and help health professionals in predicting the Diabetes disease effectively and efficiently.

**REFERENCE**

1. Kohli, Pahulpreet Singh, and Shriya Arora. "Application of machine learning in disease prediction." *2018 4th International conference on computing communication and automation (ICCCA)*. IEEE, 2018.
2. Krishnan, Santhana, and S. Geetha. "Prediction of Diabetes Disease Using Machine Learning Algorithms." *2019 1st international conference on innovations in information and communication technology (ICIICT)*. IEEE, 2019.