**SMART BRIDGE EDUCATIONAL Pvt Ltd**



**INTERNSHIP REPORT**

**ON**

**“Early Prediction of Diabetes Mellitus in Pregnant Women Using Machine Learning”**

**Bachelor of Engineering**

**In**

**Computer Science and Engineering**

**Submitted By : Ankit N R (4BD18CS006)**



**Bapuji Institute of Engineering & Technology ,Davangere- 577004**

**2021**

**CONTENTS**

**1.INTRODUCTION**

1.1 Overview 1

1.2 Purpose 1

**2 .LITERATURE SURVEY**

2.1 Existing problem 2

2.2 Proposed solution 2

**3 .THEORITICAL ANALYSIS**

3.1 Block diagram 3

3.2 Hardware / Software designing 3

**4 .EXPERIMENTAL INVESTIGATIONS** 4-5

**5 .FLOWCHART**  6

**6 .RESULT**  7-10

**7 .ADVANTAGES & DISADVANTAGES**  11

**8 .APPLICATIONS** 11

**9 .CONCLUSION**  12

**10 .FUTURE SCOPE**  13

**11 .BIBILOGRAPHY** 14

**1.INTRODUCTION**

**1.1 Overview**

Diabetes is a common chronic disease and poses a great threat to human health. The characteristic of diabetes is that the blood glucose is higher than the normal level, which is caused by defective insulin secretion or its impaired biological effects, or both. Diabetes can lead to chronic damage and dysfunction of various tissues, especially eyes, kidneys, heart, blood vessels and nerves.

With the development of living standards, diabetes is increasingly common in people’s daily life. Therefore, how to quickly and accurately diagnose and analyze diabetes is a topic worthy studying. In medicine, the diagnosis of diabetes is according to fasting blood glucose, glucose tolerance, and random blood glucose levels.

**1.2 Purpose**

People with diabetes must take responsibility for their day-to-day care.

This includes monitoring :

1. blood glucose levels
2. dietary management
3. Blood pressure
4. insulin use via injections or pump.

1

**2 .LITERATURE SURVEY**

**2.1 Existing problem**

Diabetes a non-communicable disease is leading to long-term complications and serious health problems. A report from the World Health Organisationaddresses diabetes and its complications that impact on individual physically, financially, economically over the families. The survey says about 1.2 million deaths due to the uncontrolled stage of health lead to death. About 2.2 million deaths occurred due to the risk factors of diabetes like a cardiovascular and other diseases.

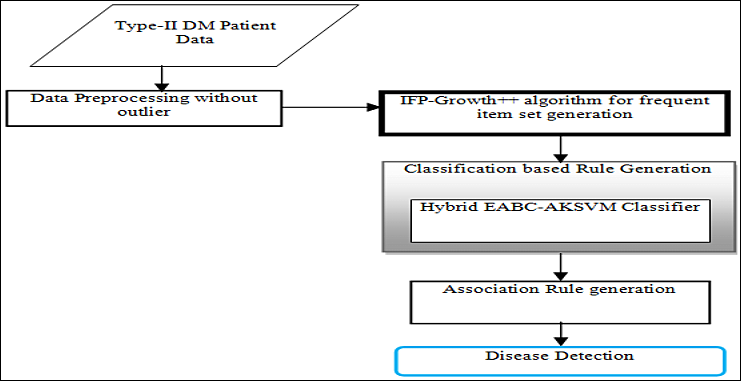
**2.2 Proposed Solution**

AI and ML supports various classifiers, decision support system is proposed that uses the AdaBoost algorithm with Decision Stump as a base classifier for classification. Moreover, Random Forest Classifier,Decision Tree Classifier,ANN have additionally executed as a base classifiers for Add a Boost calculation for exactness confirmation.

2

**3 .THEORITICAL ANALYSIS**

**3.1 Block diagram**



**3.2 Hardware / Software designing**

**HARDWARE:**

1. Processor : Intel dual core or above
2. Processor speed : 1.0 GHz or above
3. RAM : 4GB RAM or above
4. Hard Disk : 20GB hard disk or above

**SOFTWARE:**

1. Language : Python
2. User Interface Design: HTML
3. Web Browser : Google Chrome

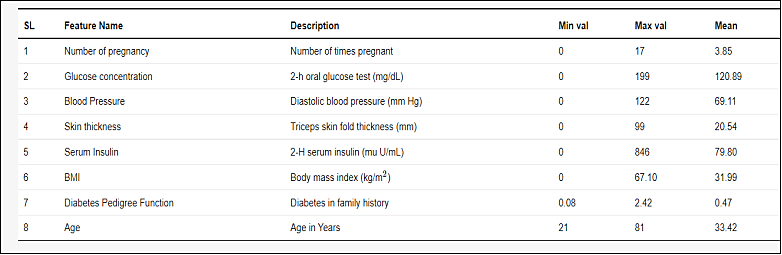
3

**4 .EXPERIMENTAL INVESTIGATIONS**

**Data set:**

The PIMA Indians Diabetes dataset obtained from Kaggle , has been originally collected to test the ML model and predict diabetes patients.

Both datasets consist of female patients which is why the number of times of pregnancy is one of the most important features for our analysis. We have evaluated several Machine Learning classifiers on these particular datasets to predict whether a female patient is suffering from diabetes or not.

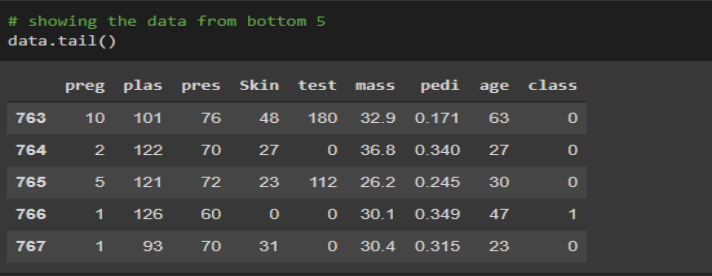
**Analysis:**

1. head() method is used to return top n (5 by default) rows of a DataFrame or series.

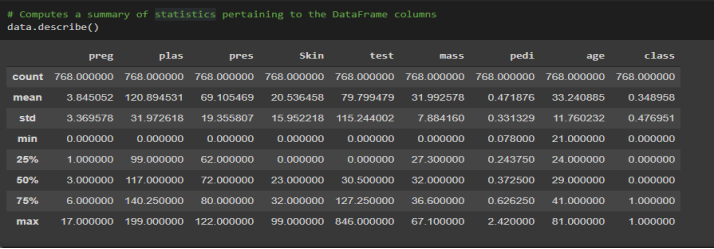


4

1. tail() method is used to return bottom n (5 by default) rows of a DataFrame or series.

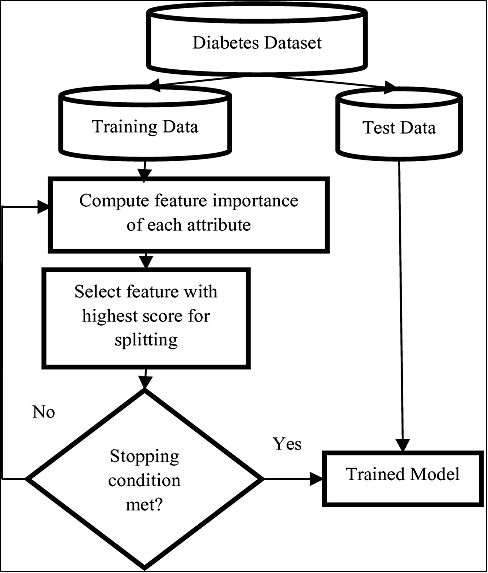


1. describe() method computes a summary of statistics like count, mean, standard deviation, min, max, and quartile values.



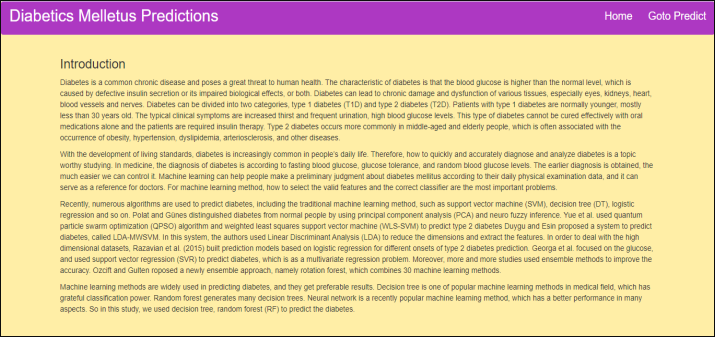
5

**5 .FLOWCHART**

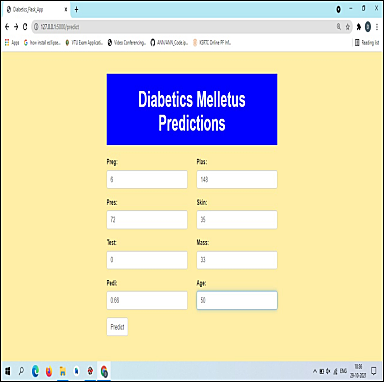


6

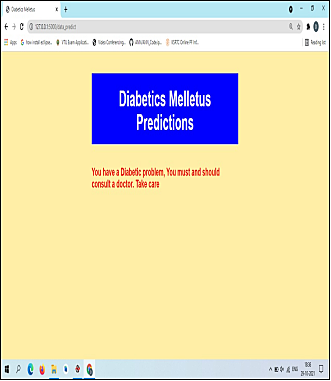
**6 .RESULT**



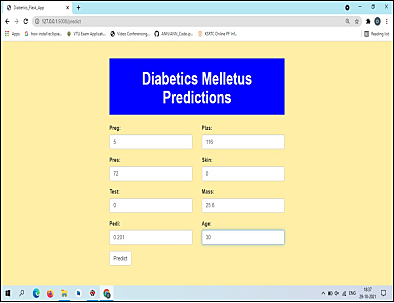
Output 1:



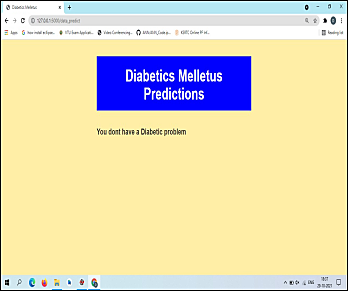
7



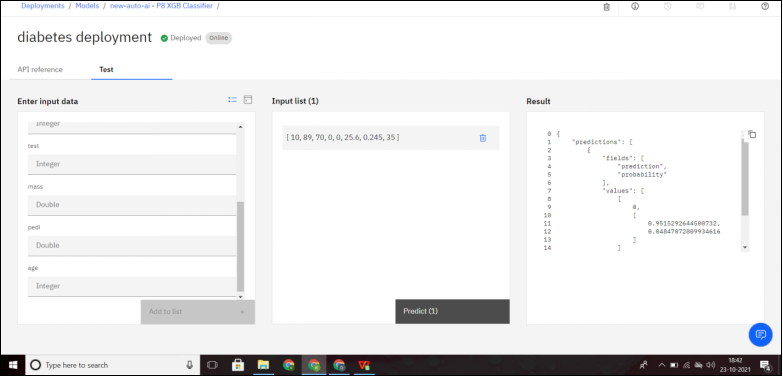
Output 2:



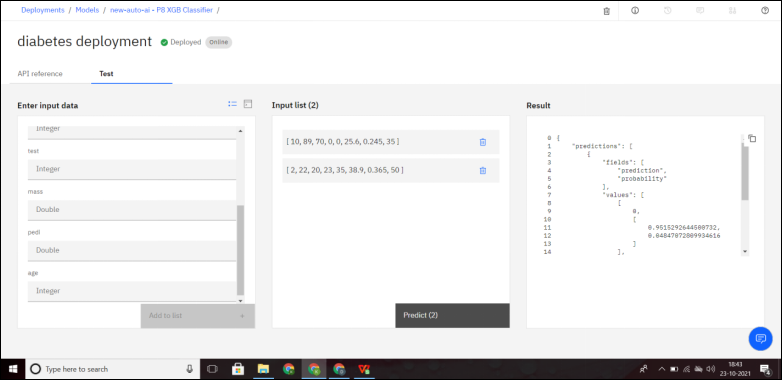
8



In Watson Stdio:



9



10

**7 .ADVANTAGES & DISADVANTAGES**

**Advantages:**

1. Easily identifies trends and patterns.
2. No human intervention needed
3. Continuous Improvement.
4. Handling multidimensional and multivariety data
5. Wide applications

**Disadvantages:**

1. High error susceptibility.
2. Data acquisition.
3. Interpretation of results.
4. Time and resources.

**8 .APPLICATIONS**

1. Medical Field
2. Research field

11

**9 .CONCLUSION**

Machine learning techniques are valuable in disease diagnosis. The capability to predict diabetes early, assumes a vital role for the patient's appropriate treatment procedure. In this a few existing classification methods for medical diagnosis of diabetes patients have been discussed on the basis of accuracy. An classification problem has been detected in the expressions of accuracy. Three machine learning techniques were applied on the Pima Indians diabetes dataset, as well as trained and validated against a test dataset. The results of our model implementations have shown that Random Forest Classifier outperforms the other models. Using association rule mining, the results have shown that there is a strong association of BMI and glucose with diabetes.

12

**10 .FUTURE SCOPE**

The limitation of this study is that a structured dataset has been selected but in the future, unstructured data will also be considered, and these methods will be applied to other medical domains for prediction, such as for different types of cancer, psoriasis, and Parkinson's disease. Other attributes including physical inactivity, family history of diabetes, and smoking habit, are also planned to be considered in the future for the diagnosis of diabetes.

13

**11 .BIBILOGRAPHY**

1. <https://www.kaggle.com/uciml/pima-indians-diabetes-database>
2. <https://www.w3schools.com/html/>

14