****

**Green University of Bangladesh**

**Department of Computer Science and Engineering (CSE)**

**Faculty of Science and Engineering**

**Semester: (Summer, Year:2022), B.Sc. in CSE (Day)**

**Course Title: Data Mining Lab**

**Course Code: CSE- 424 Section: 191D4**

**Lab Project Name: Diabetes Prediction Base on Weka and Python.**

**Student Details**

|  |  |  |
| --- | --- | --- |
| **Name** | | **ID** |
| **1.** | **MD Shahrior Shawon Prio** | 192002032 |
| **2.** | **MD Masum Billah** | 191002053 |
| **3.** |  |  |

**Submission Date : 11/09/22**

**Course Teacher’s Name : Arpita Das, Lecturer, Dept. of CSE**

**[For Teachers use only: Don’t Write Anything inside this box]**

|  |
| --- |
| **Lab Project Status**  **Marks: ………………………………… Signature: .....................**  **Comments: .............................................. Date: ..............................** |

Table of Contents

Chapter 1 Introduction 3

1.1 Introduction 3

1.2 Design Goals/Objective 3

Chapter 2 Design/Development/Implementation of the Project 4

2.1 Section (Choose the name of this section as appropriate with your project) 4

2.2 Section (Choose the name of this section as appropriate with your project) 4

2.2.1 Subsection 4

Chapter 3 Performance Evaluation 5

3.1 Simulation Environment/ Simulation Procedure 5

3.2 Results and Discussions 5

Chapter 4 Conclusion 6

4.1 Introduction 6

4.1 Practical Implications 6

4.2 Scope of Future Work 6

References 7

# Chapter 1 Introduction

## Introduction

## This project is about to train dataset into weka and observing the changes on accuracy by applying Decision Tree algorithm and Artificial neural network.

## Design Goals/Objective

In this project our general objective is to find the variation of accuracy on Decision Tree algorithm and Artificial neural network and draw 3 graphs.

By 3 differences:

1. Fold number vs accuracy

2. Hidden Layers vs accuracy

3. Training Time vs accuracy

So, we made a project using tools, and the practices that we learned from the lab of this course.

# Chapter 2

# Design/Development/Implementation of the Project

## 2.1.1 Tools:

## Weka 3.8.5 WEKA

## Net Beans

## Google Colab

## Techniques:

## Java

## Python

## 2.2.2 Methodology:

## 

# Chapter 3 Performance Evaluation

## Simulation Environment/ Simulation Procedure

Screenshot:

**Procedure (j48 for Decision Tree and Multilayer Perception in Weka):**

**Graphical user interface, text

Description automatically generated** **Graphical user interface, text, application

Description automatically generated**

**Visualizing Tree:**

**A picture containing text

Description automatically generated**

**Visualizing Neural Network:**

**Chart

Description automatically generated**

**Now setting up five different fold number to determine accuracy (10):Text

Description automatically generated with low confidence**

**Now setting up five different fold number to determine accuracy (20):**

**Graphical user interface, text

Description automatically generated**

**Now setting up five different fold number to determine accuracy (50):**

**Text

Description automatically generated**

**Now setting up five different fold number to determine accuracy (500):Graphical user interface, text

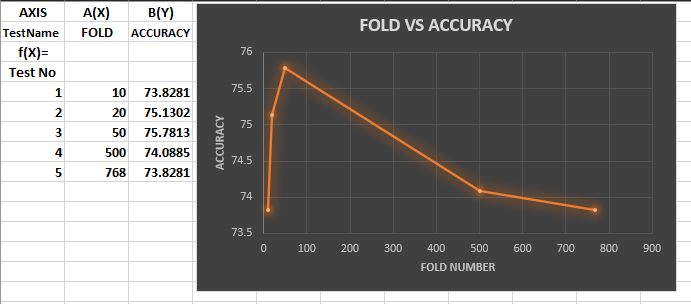
Description automatically generated**

**Now setting up five different fold number to determine accuracy (768):**

**Text

Description automatically generated with medium confidence**

**Graph 1: Fold vs Accuracy:**

****

**Now setting up five different Hidden Layers to determine accuracy (2):**

**Graphical user interface, text, application

Description automatically generated**

**Now setting up five different Hidden Layers to determine accuracy (3):**

**Graphical user interface, text, application

Description automatically generated**

**Now setting up five different Hidden Layers to determine accuracy (4):**

**Graphical user interface, text, application

Description automatically generated**

**Now setting up five different Hidden Layers to determine accuracy (5):**

**Graphical user interface, text, application

Description automatically generated**

**Now setting up five different Hidden Layers to determine accuracy (6):**

**Graphical user interface, text, application

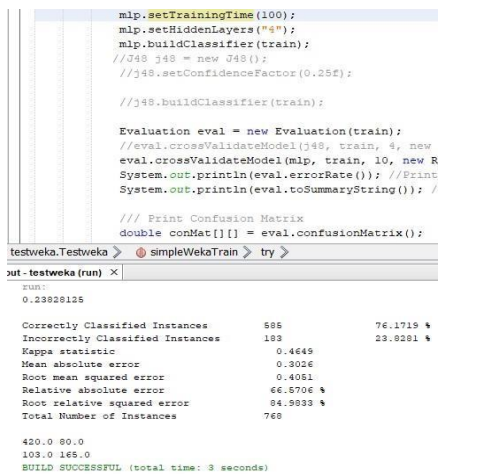
Description automatically generated**

**Graph 2: Hidden vs Accuracy:**

**Chart, line chart

Description automatically generated**

**Now setting up five different Training time to determine accuracy (100):**

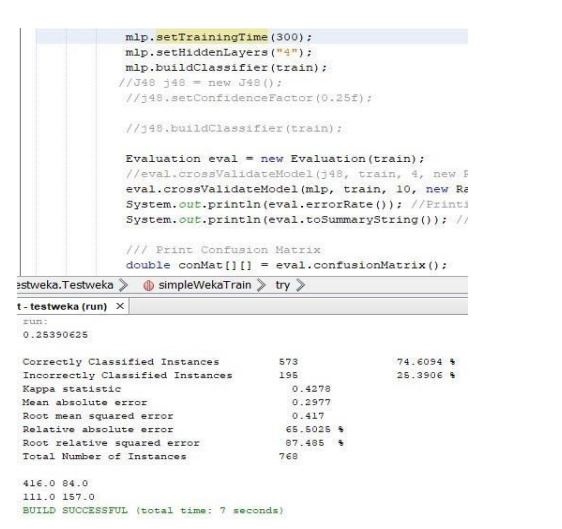
****

**Now setting up five different Training time to determine accuracy (200):**

**Graphical user interface, text, application

Description automatically generated**

**Now setting up five different Training time to determine accuracy (300):**

****

**Now setting up five different Training time to determine accuracy (400):**

**Text

Description automatically generated with medium confidence**

**Now setting up five different Training time to determine accuracy (500):**

**Graphical user interface, text, application

Description automatically generated**

**Graph 3: TrainingTime vs Accuracy:**

**Chart, line chart

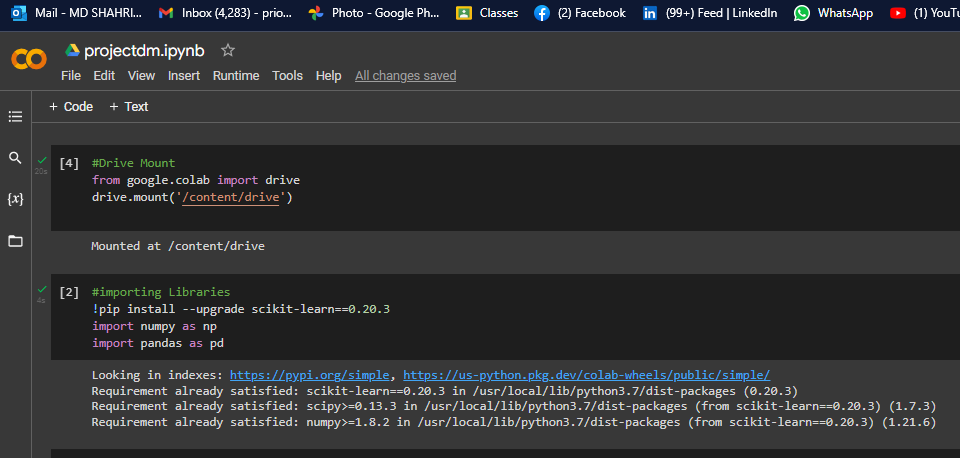
Description automatically generated**

Here, we can see accuracy is gradually decreasing on increasing the Training time.

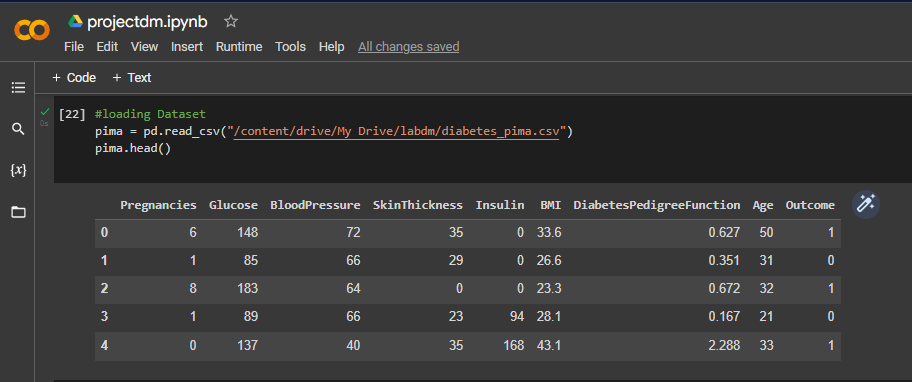
## Results and Discussions

* + 1. **Python Codes and the outputs:**

**Drive Mounting, importing libraries, training dataset, feature and target value:**

****

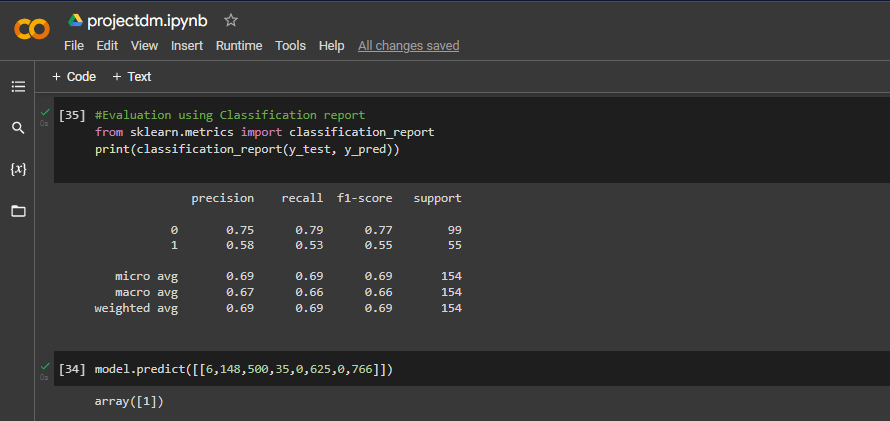
**Loading Dataset:**

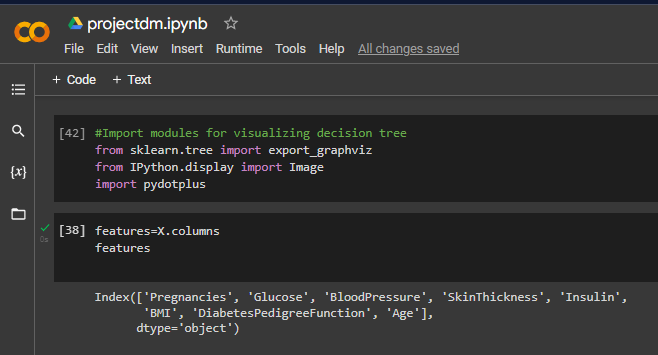
**A screenshot of a computer

Description automatically generatedText

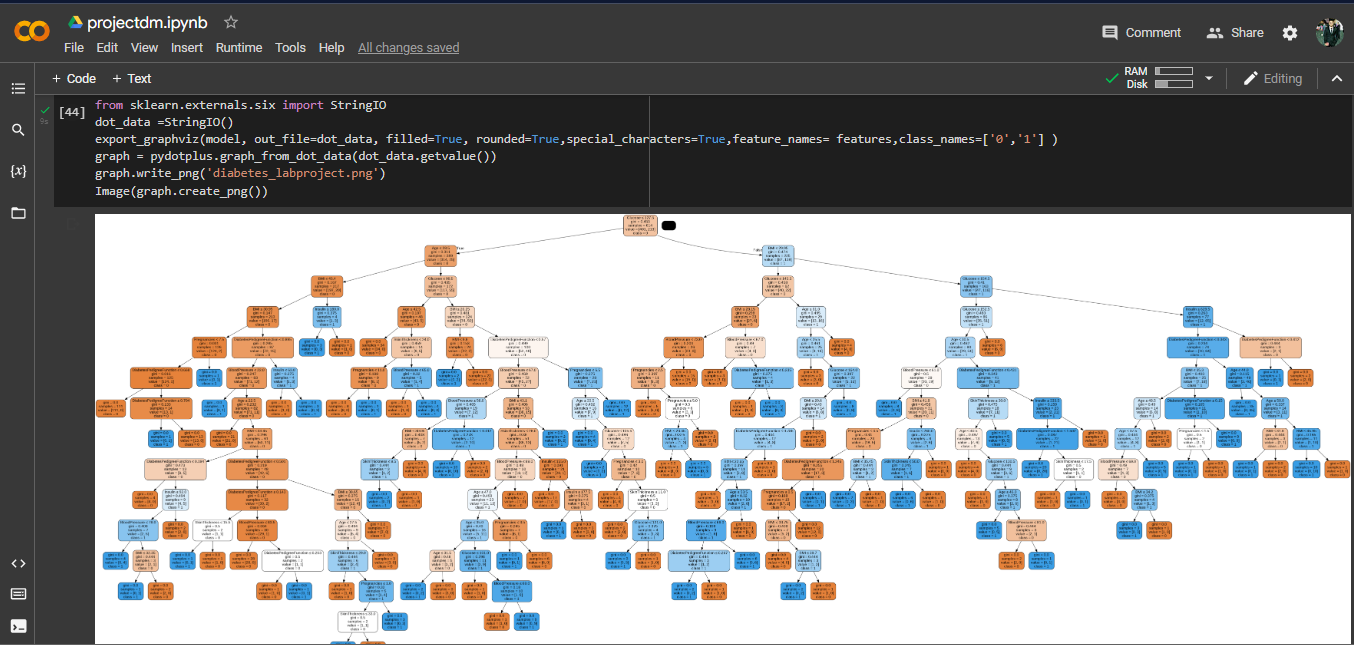
Description automatically generatedText

Description automatically generated**

****

****

**Let’s see the visualizations:**

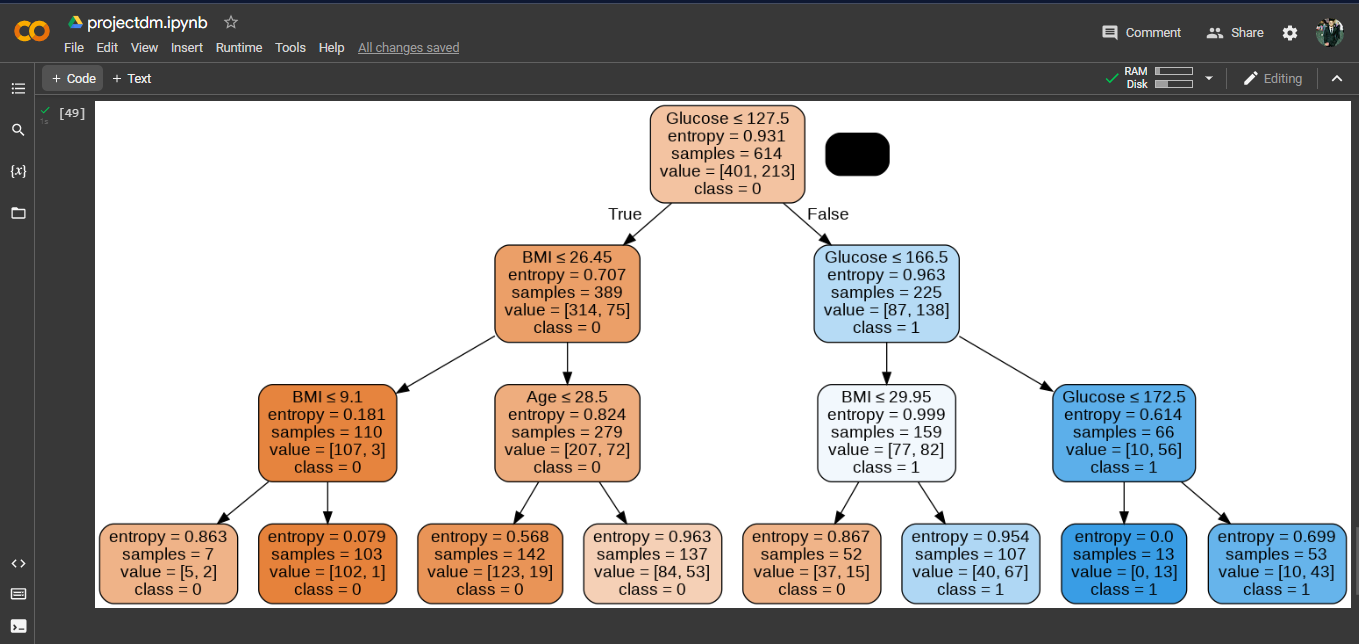
****

**Creating classifier object, train it, predict response then see the accuracy:**

**Text

Description automatically generated**

**Let’s see a clear visualization of decision tree:**

****

# Chapter 4 Conclusion

## Discussion

## From the output of decision tree visualization, we can see the condition of

## true and false. Here, glucose, BMI, age, entropy, samples value is limited

## and condition will true and class must = 0. Then perfectly, we define that people have no diabetes. So, they are always diabetes free. On the other

## hand, people have diabetes, and they are not diabetes free.

Finally, we are finish all work this project.

## Scope of Future Work

* + - When we were working on this project, we faced many problems in colab.
    - In future, we are trying to solve all limitation and problem with advance decision tree algorithm.

# References

1. Most of them are from lab practices and lab manuals.
2. The practices we did from the labs.
3. [www.youtube.com](http://www.youtube.com)
4. [www.stackoverflaw.com](http://www.stackoverflaw.com)
5. <https://github.com/>