DIABETES PREDICTION

USING PYTHON

OBJECTIVES OF THE PROJECT

- The motive is to **design a model** which can prognosticate the likelihood of diabetes in patients with maximum accuracy.
- Using machine learning techniques for identifying diabetes at early stage and with proper treatment which can be controlled.
- **The main purpose of using machine learning was it helps to build the system that should learn from previous experiences and complete the tasks by itself without any need of external instruction.**

OVERVIEW OF THE PROJECT

- SVM machine learning algorithm is used in this system to detect diabetes at an early stage.
- Machine learning model will predict output for a particular given input.
- This proposed method can predict the diabetes with higher accuracy levels than the traditional methods.
- The output can also be represented in diagrammatical and graphical format.
- Results obtained predict whether a person has diabetes or not with high level of accuracy.
- The performances of algorithms can be evaluated by various measures like Precision, Accuracy, and Recall and the result is displayed in the **Confusion Matrix.**

DATASET USED FOR THE PROJECT

The dataset is been obtained from **Kaggle.**

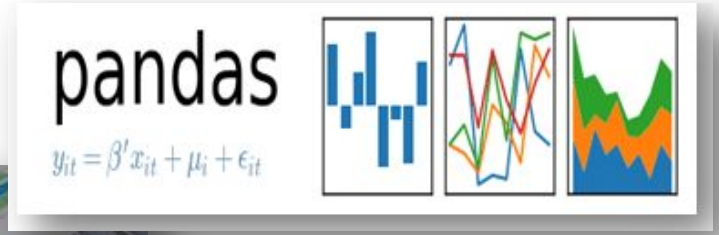
- PREG: This column indicates how many times a person is pregnant.
- PLAS: This indicates plasma glucose concentration at 2 h in an oral glucose tolerance test
- PRES: This shows diastolic blood pressure.
- SKIN: This indicates thickness of skin at triceps.
- INSULIN: This demonstrates insulin level.
- BMI: It demonstrates body mass index which is ratio of weight and height.
- PEDI: It demonstrates how much probability a person can inherit diabetes from ancestors.
- Age: It provides or shows age of the person.
- Class: It is a variable which contains only 0 or 1.

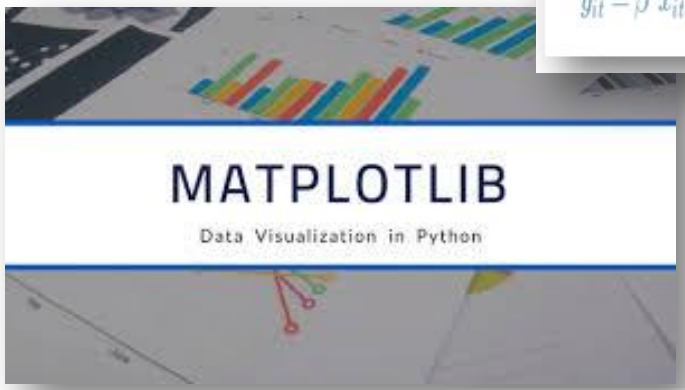
1 indicates person having diabetes and 0 indicates person not having diabetes.

No.	Attribute
1	Number of times pregnant
2	Plasma glucose concentration in an oral glucose tolerance test
3	Diastolic blood pressure (mm/Hg)
4	Triceps skin fold thickness (mm)
5	2-hour serum insulin (μU/ml)
6	Body mass index (kg/m^2)
7	Diabetes Pedigree function
8	Age (years)

Modules used in project

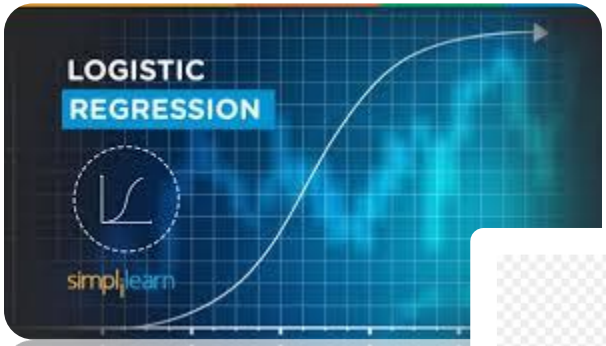
- Warnings
- Pandas
- Matplotlib
- Seaborn
- Numpy
- Sklearn
- Tkinter

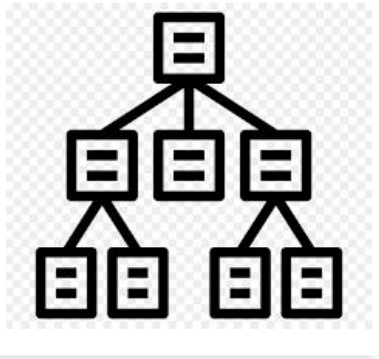




ALGORITHMS USED IN THE PROJECT

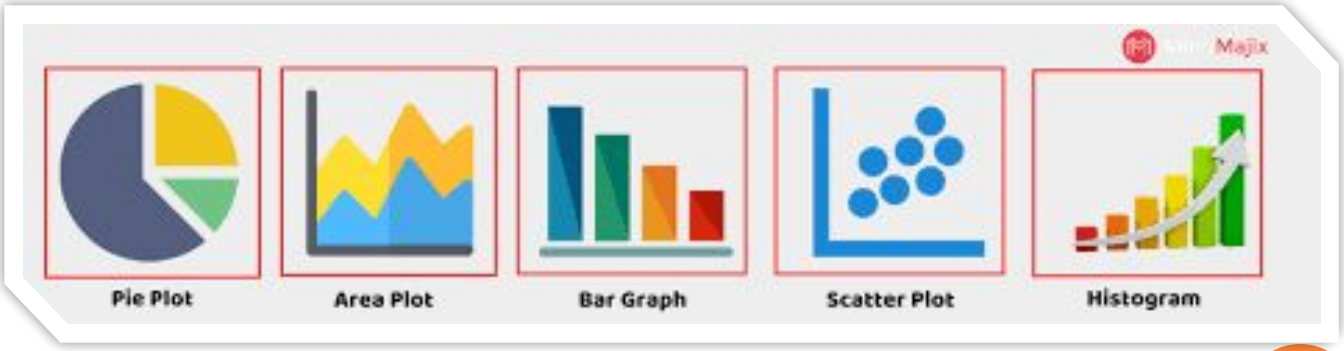
- Decision tree classifier
- Random forest classifier
- Gradient boosting classifier
- Logistics regression
- Min max scalar
- Sv classifier(prediction)





VISUALISATION(GRAPHS)

- Histogram
- Bar graph
- Scatter plot
- Distplot
- Boxplot
- Heatmap
- Decision tree



System configuration

Hardware Requirements:

- Processor: Any Update Processor
- Ram : Min 4 GB
- Hard Disk : 100 GB

Software Requirements:

- Operating System: Any Windows Family or Linux
- Technology : Python 3.5 and above versions
- User Interface : Tkinter Module in Python

Development Tools

- Anaconda(Spyder)
- Python using terminal in Linux

FEASIBILITY

- **Technical Feasibility**

The platform will be portable enough on any machines. Python is a basic requirement to build the desired system.

- **Operational Feasibility**

The system is usable and does not require any extra knowledge or a technical person to operate it.

- **Economic Feasibility**

No extra cost required to be spent to develop this system.

- **Time Feasibility**

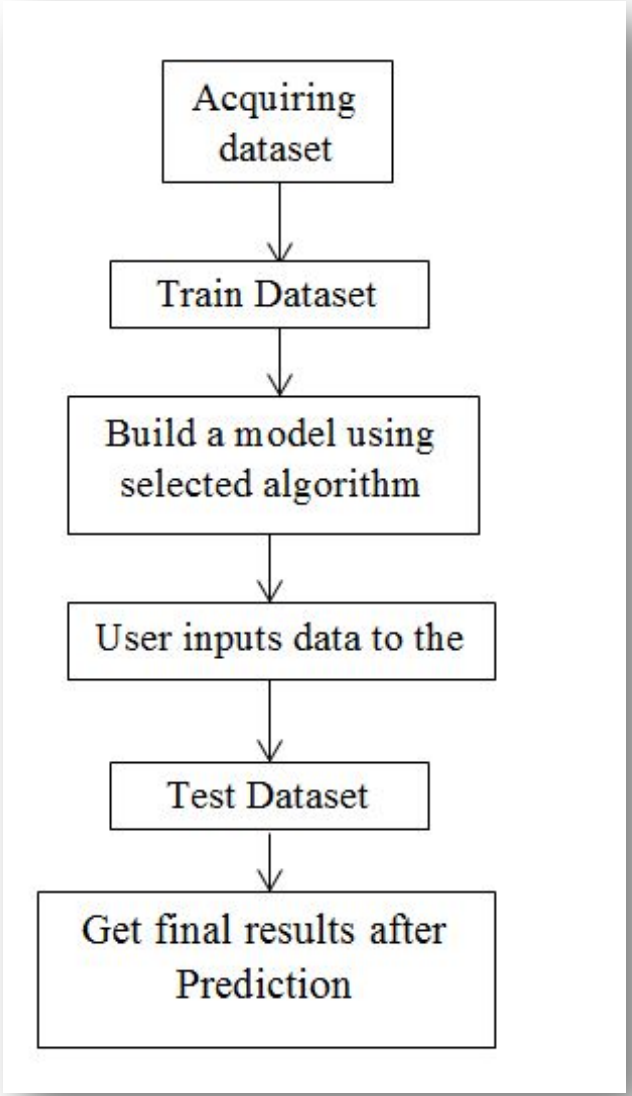
Design to be developed in a period of 2 months.

Methodology:

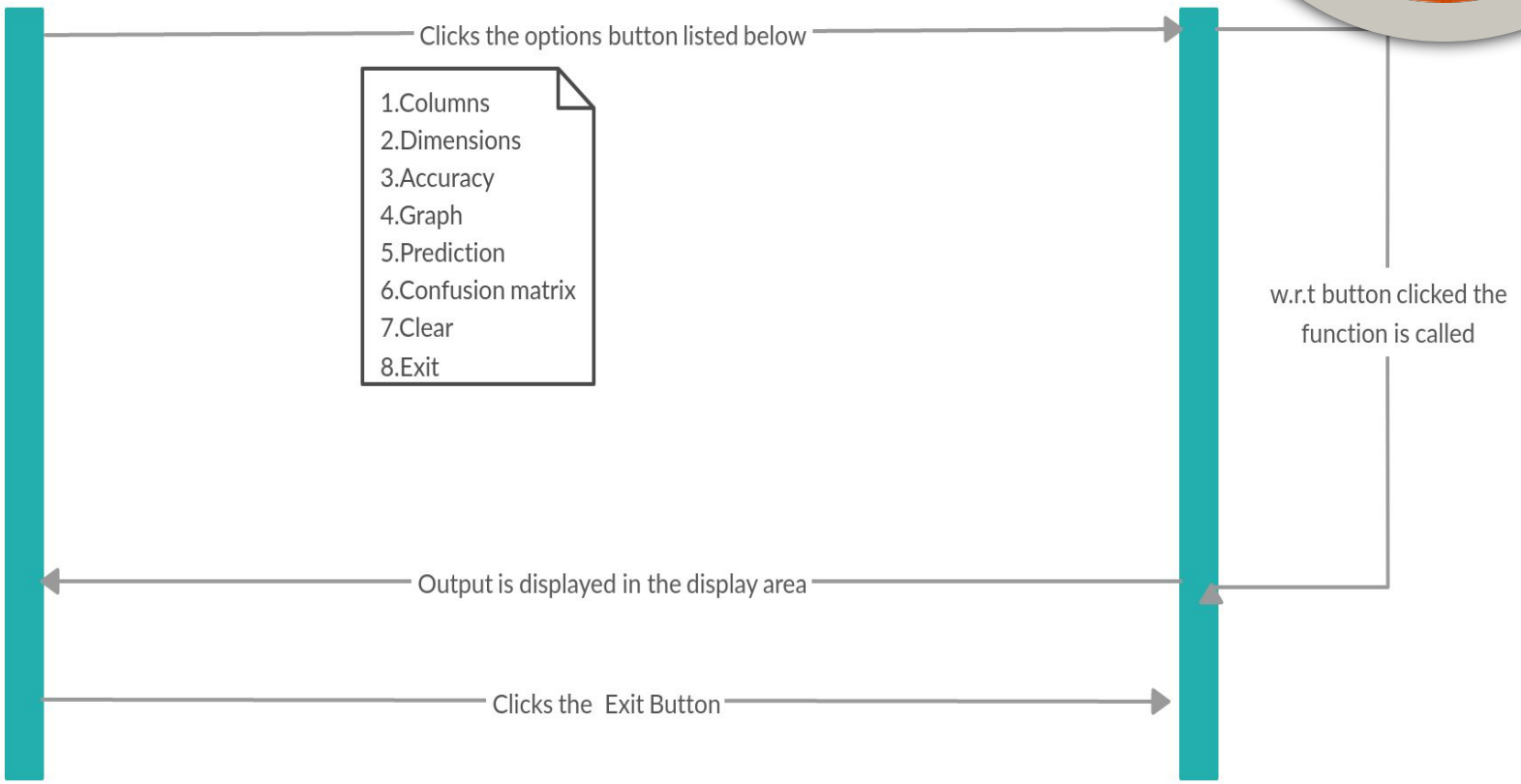
This section explains the overall design of the system and what is the process it has followed in order to get the prediction.

Procedure:

- Load previous data sets to the system .
- Build a model using the algorithm selected and train the data set
- Then User inputs data to the system in order to diagnose whether he has the disease or not.
- Test the data set using the model.
- Finally, get the predicted diagnostic result also the visualisation(if required) of result can be seen accordingly.

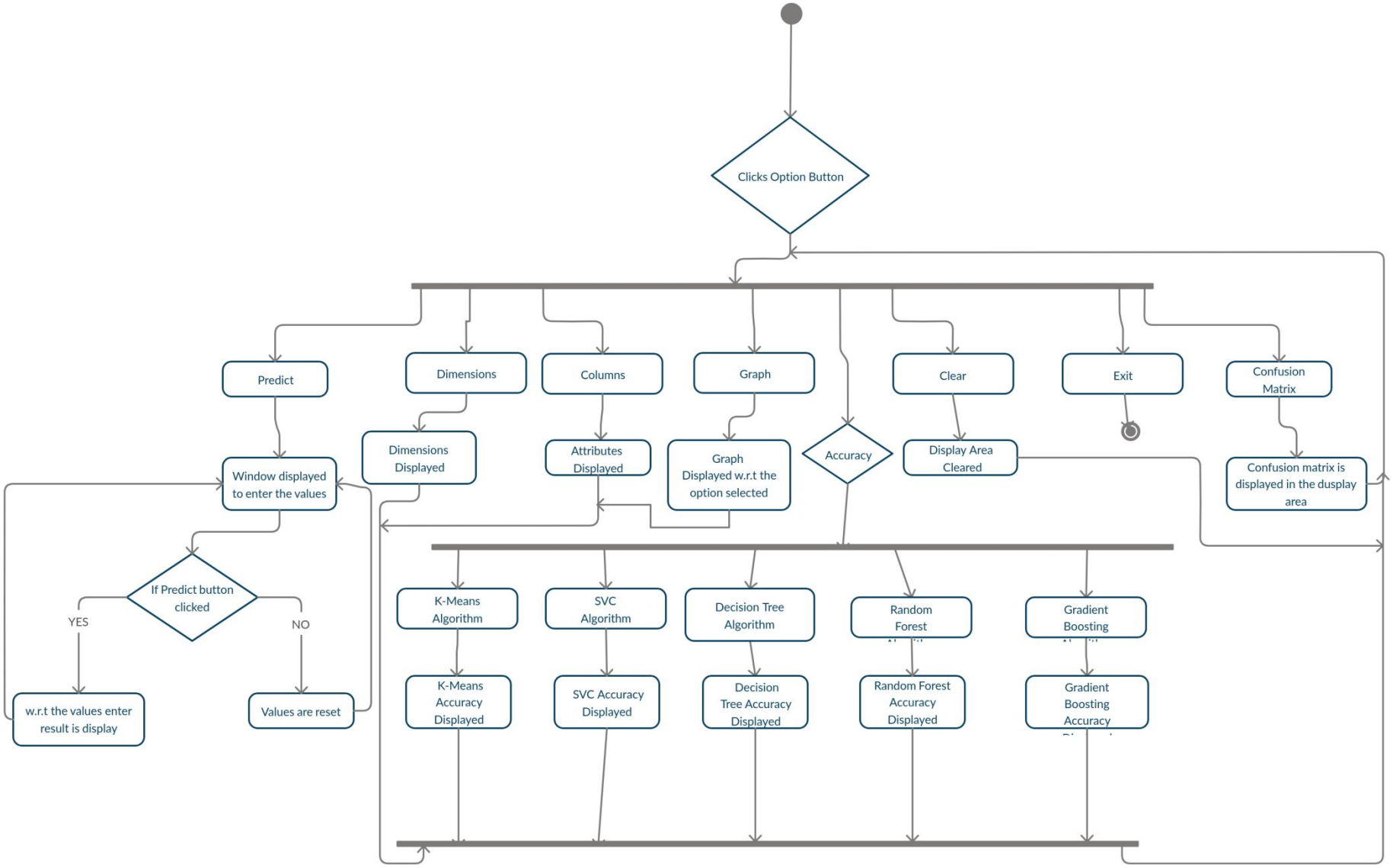


Diagrams of the project



SEQUENCE DIAGRAM

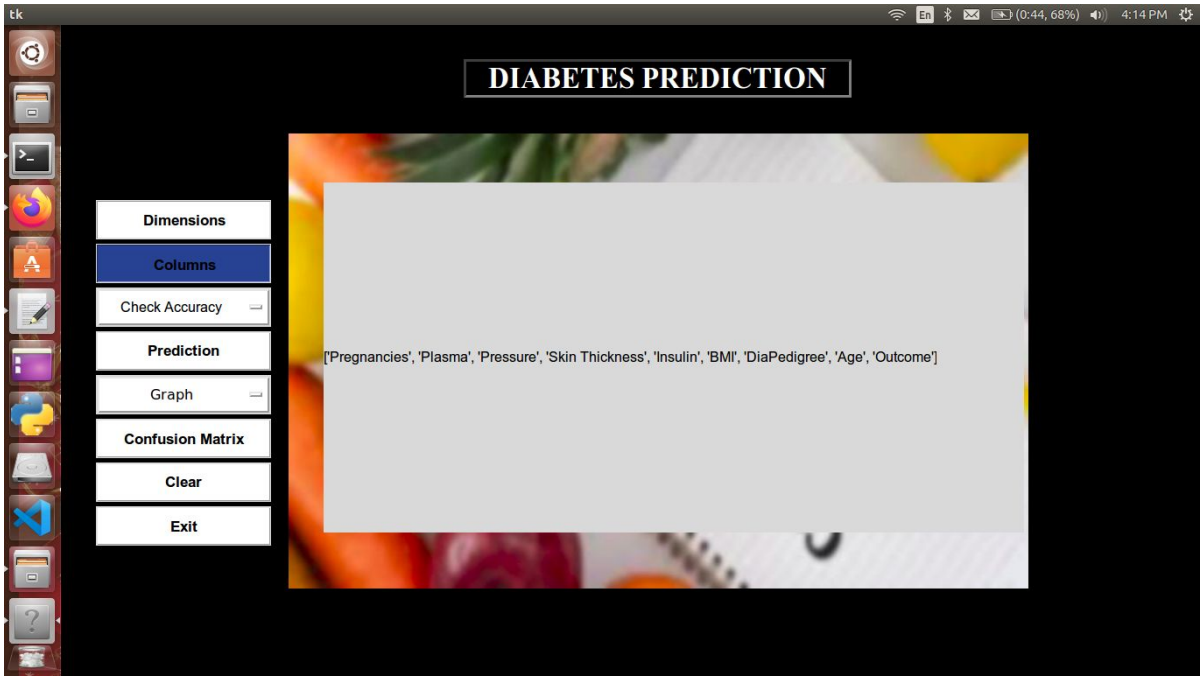
USE CASE DIAGRAM



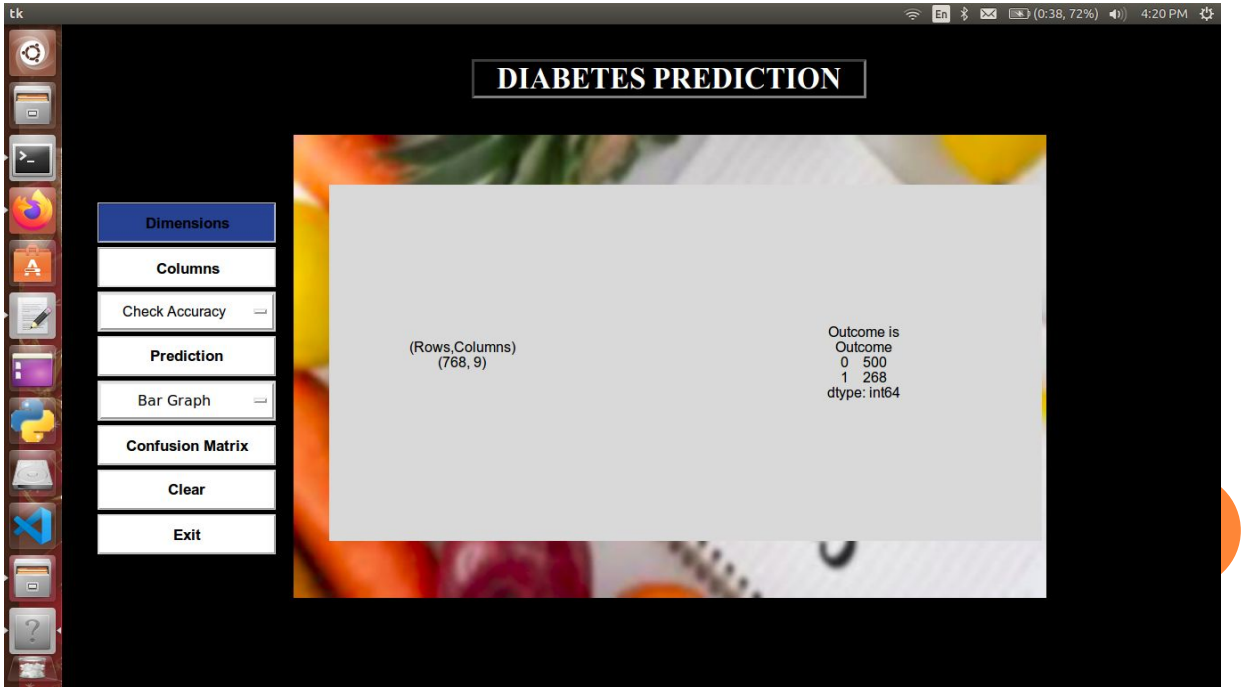
ACTIVITY DIAGRAM

IMPLEMENTATION : SCREENSHOTS

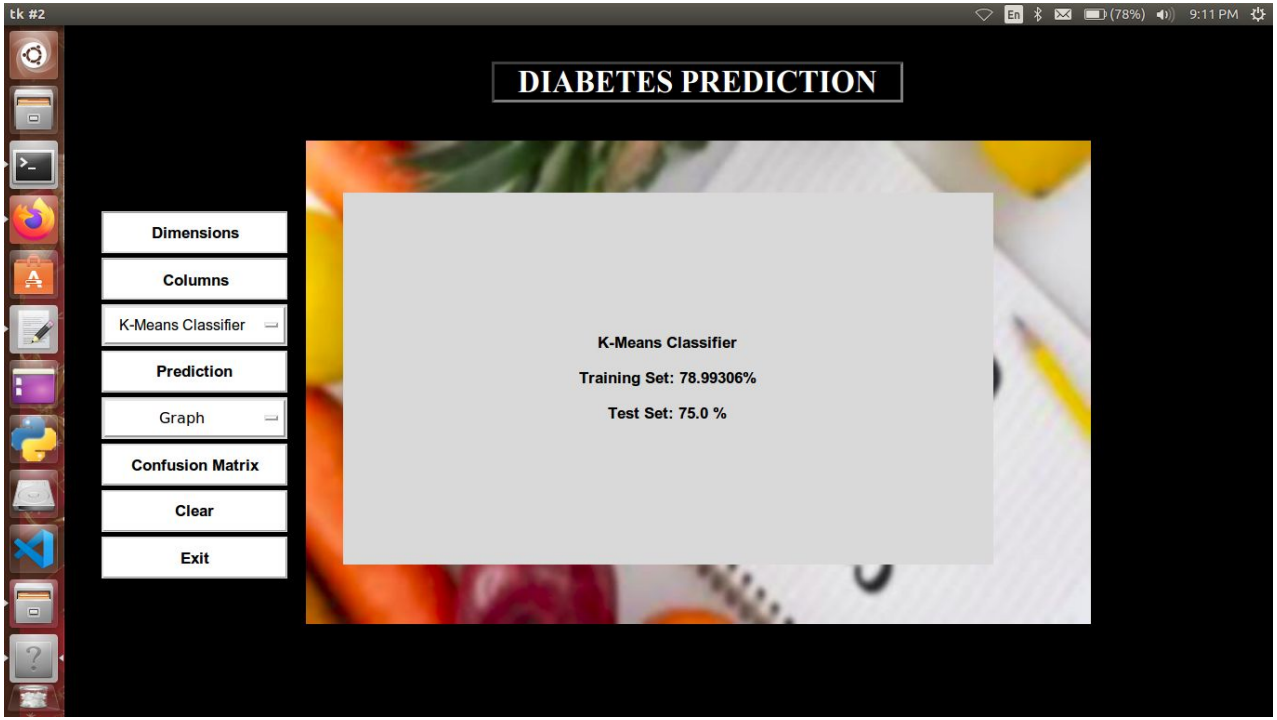
FIRST(ACTUAL) SCREEN



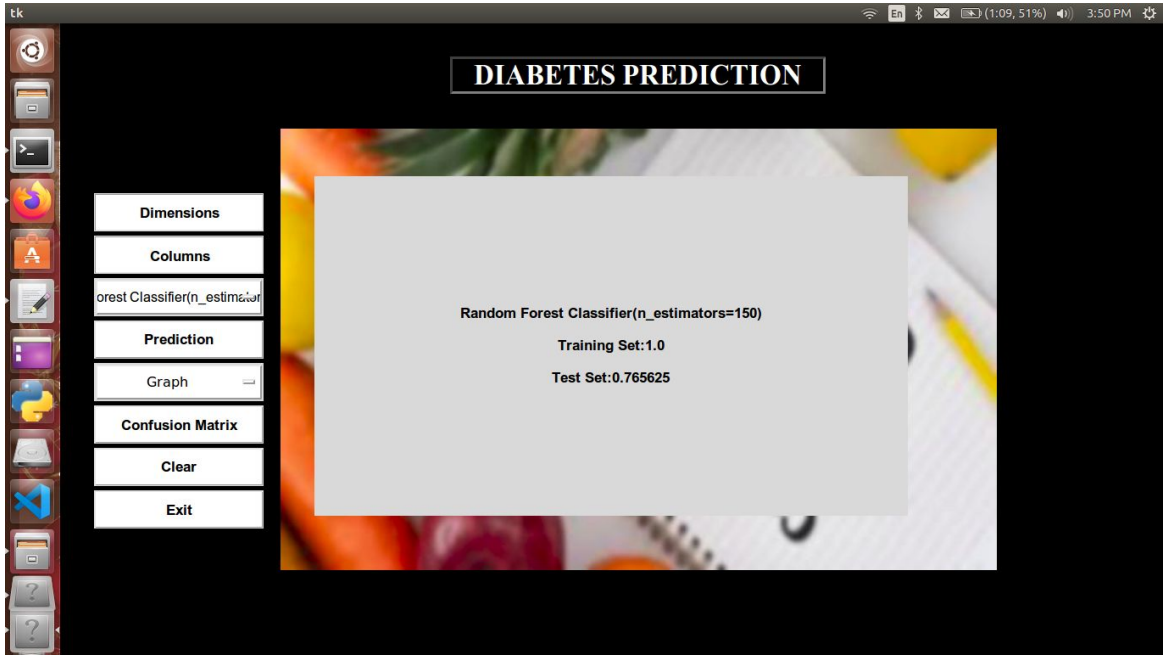
COLUMNS



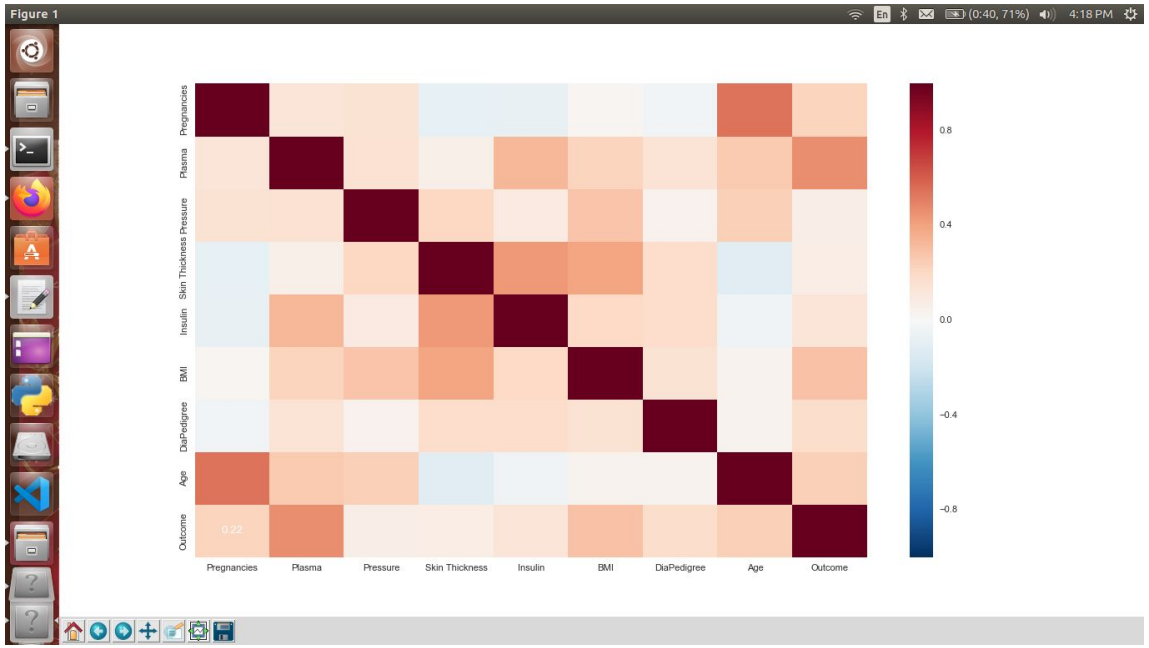
DIMENSIONS



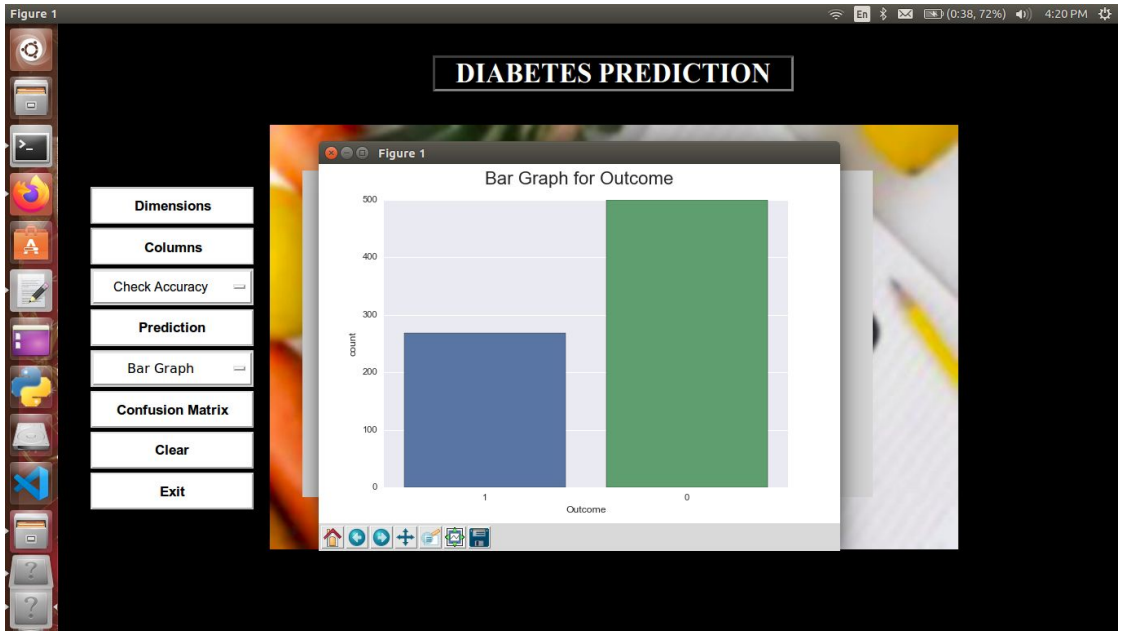
K-MEANS ALGORITHM



RANDOM FOREST CLASSIFIER

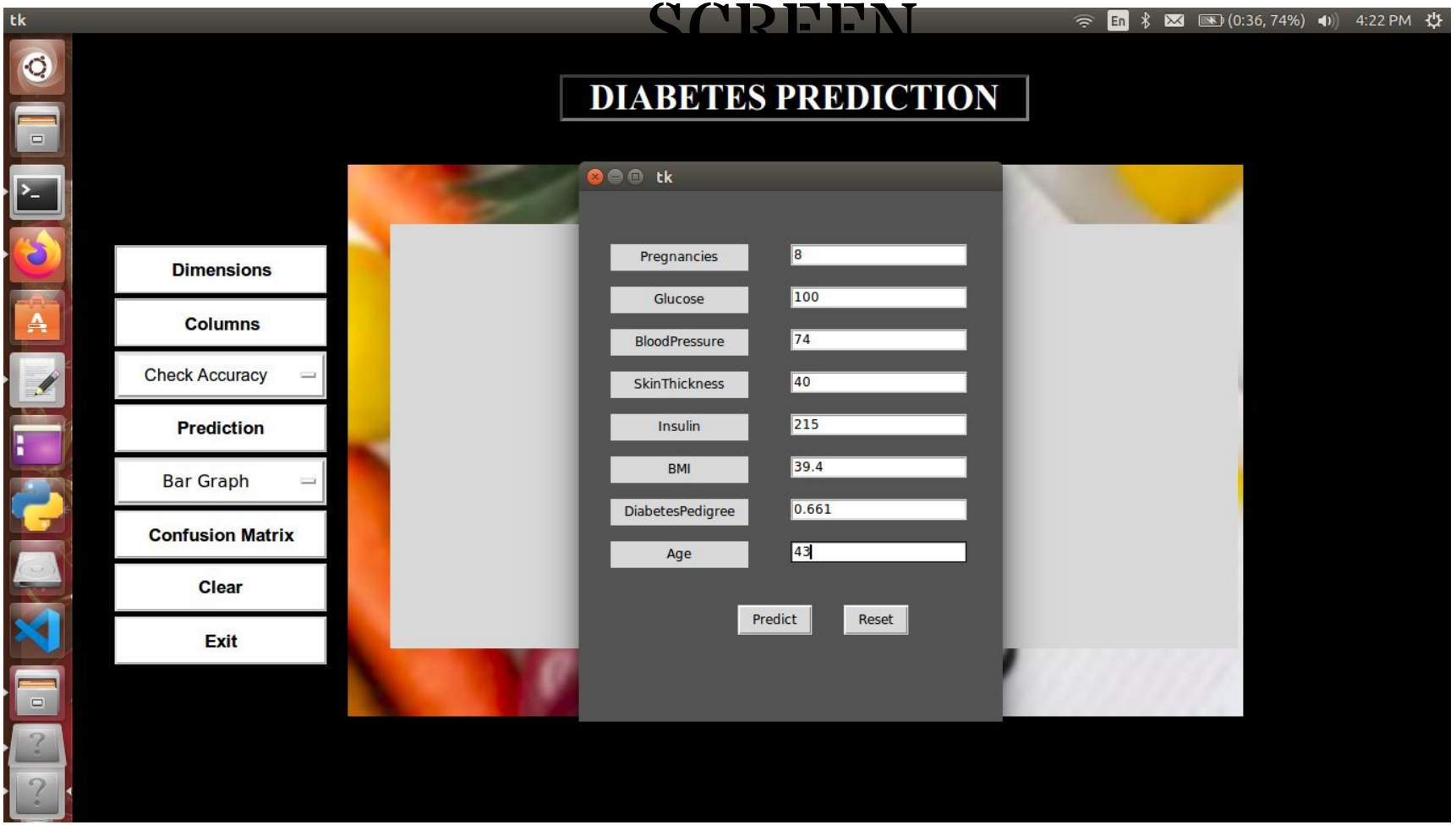


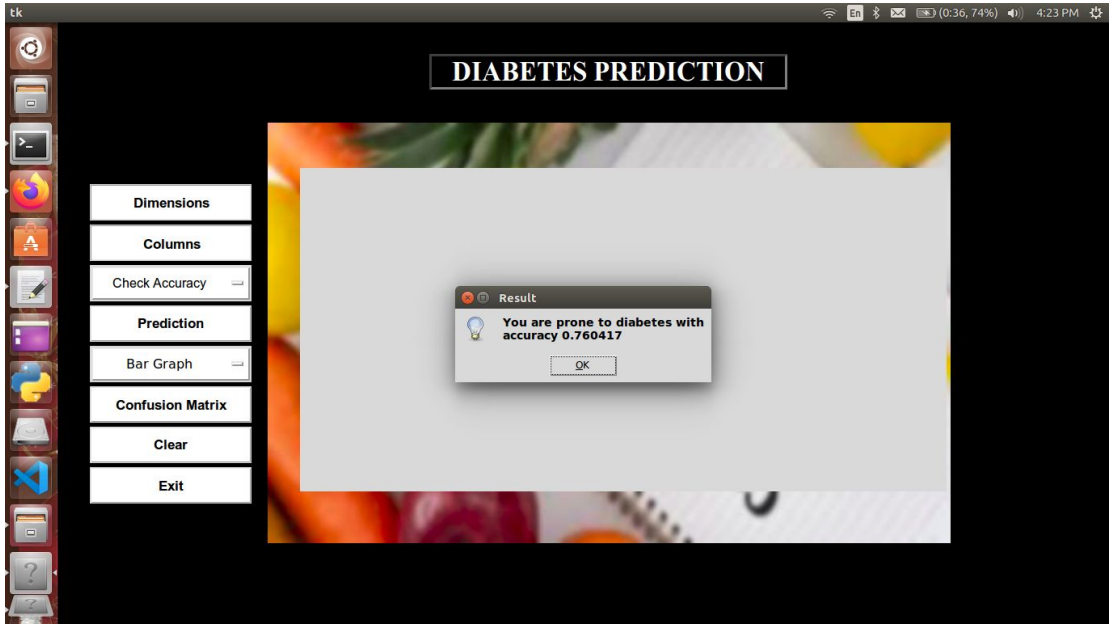
HEAT MAP



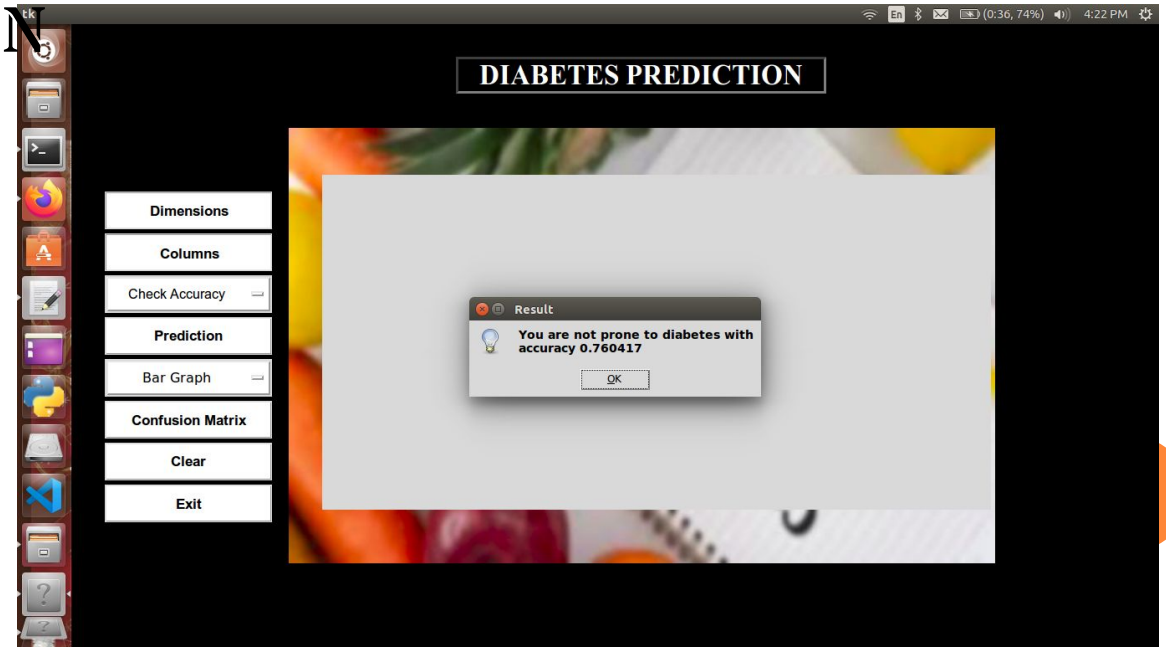
BAR GRAPH

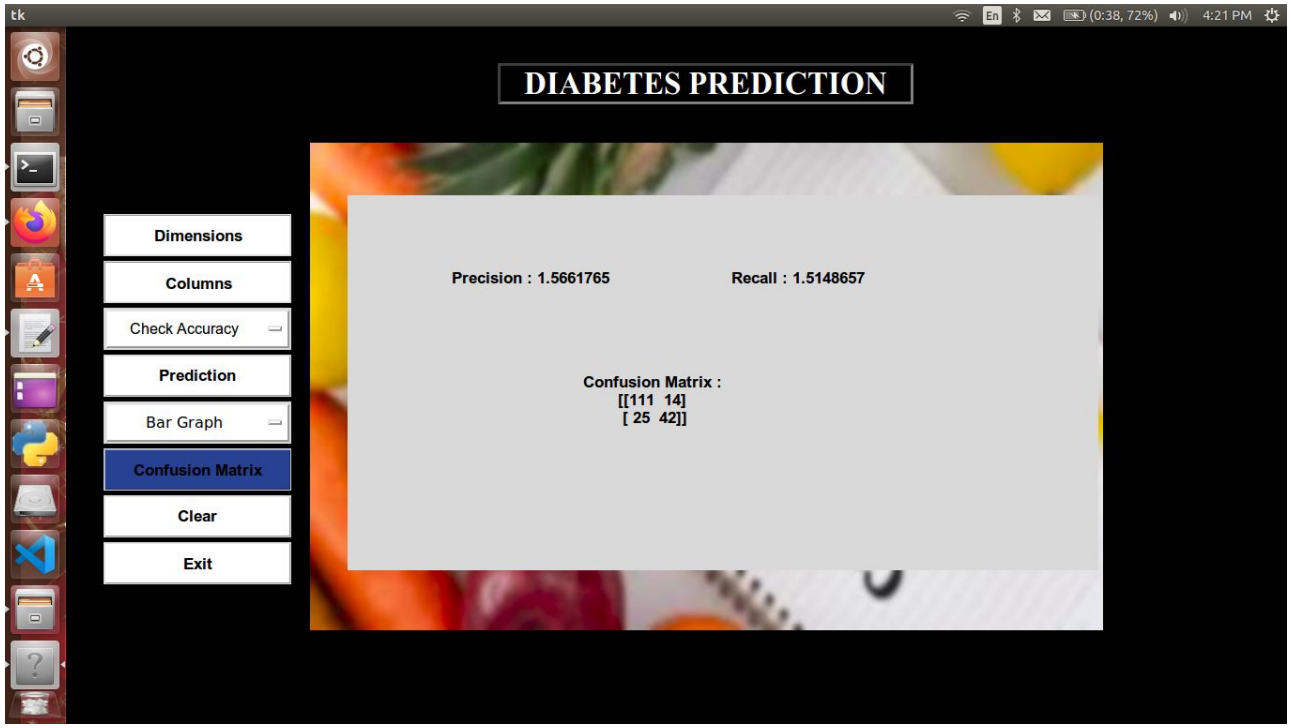
PREDICTION SCREEN



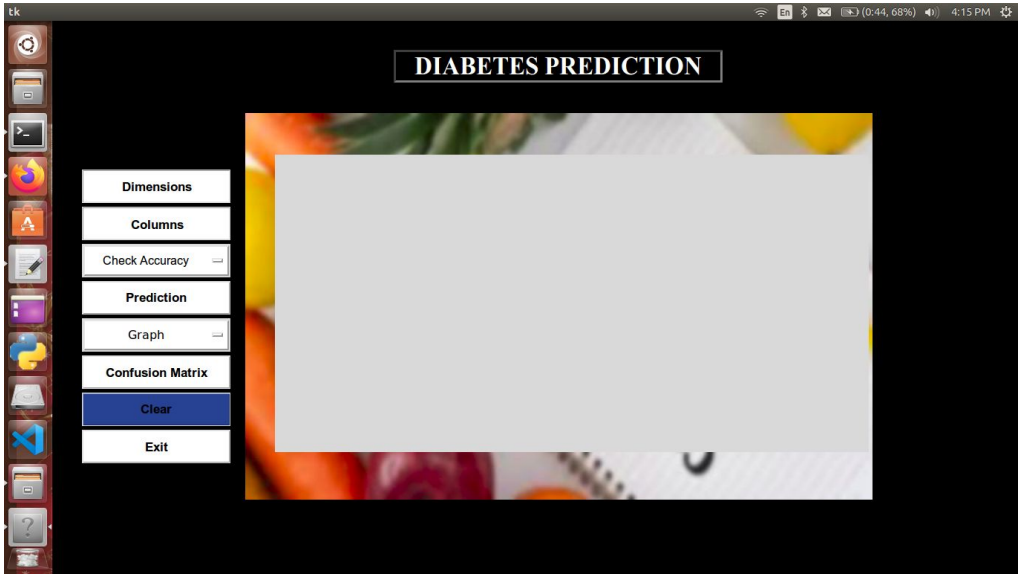


OUTCOMES OF THE PREDICT FUNCTION





CONFUSION MATRIX



CLEAR BUTTON

TESTING

- Testing is the process of evaluating a system or its component(s) with the intent to find whether it satisfies the specified requirements or not.
- Testing is executing a system in order to identify any gaps, errors, or missing requirements in contrary to the actual requirements.
- The usage of the word "**testing** " in relation to **Machine Learning** models is primarily used for **testing** the model performance in terms of accuracy/precision of the model.

Test cases for the system:

Input	Expected O/P	Actual O/P
Dimensions	No of rows and coulmns displayed	Yes
Columns	Names of the columns	Yes
Multiple algorithms • Random Forest Classifier • SV Classifier • Logistic Regression	Accuracies of the algorithms are displayed accordingly	Yes
Prediction	Output is shown whether a Person is pruned to diabetes or no	Yes/No Depending upon the values given as input.
Graph • Histogram • Bar Graph • Scatter Plot • Box Plot	The output is visualised according to the Option selected.	Yes
Confusion Matrix	The confusion matrix is displayed.	Yes

Drawbacks and limitations

- The system does not provide 100% accuracy.
- The current system can be enhanced to predict about other diseases too for now it is restricted for diabetes only.
- A perfectly developed model for diagnosis of diabetes will require more training data for creation and testing whereas our dataset comprises of smaller amount of data.
- The current system predicts the output but cannot specify a particular attribute due to which the patient is pruned to diabetes.

Conclusion

- The automatic diagnosis of diabetes is an important real-world medical problem. Detection of diabetes in its early stages is the key for treatment.
- Results show the effectiveness of the proposed model. The performance of the techniques was investigated for the diabetes diagnosis problem.
- Results demonstrate the adequacy of the proposed model.

FUTURE ENHANCEMENT

- In future it is planned to gather the information from different locales over the world
- Try to make a more precise and general prescient model for diabetes conclusion.
- Future study will likewise focus on gathering information from a later time period.
- Also discover new potential prognostic elements to be incorporated.
- The work can be extended and improved for the automation of diabetes analysis.

Presented By:

- Mariam Andagi(524))
- Pranoti Gaikwad(536)
- Manasi Basale(546)

Thank You