A Project/Dissertation ETE Viva Report

On

FITNESSO: Fitness Score Predictor and AI Dietician

Submitted in partial fulfillment of the
Requirement for the award of the degree of

Bachelors of Technology in Computer Science and Engineering

Under The Supervision of
Dr. Arvind Kumar
(Professor)

Submitted By

Chirag Kaushik and Sakshi Gaur
(19021011874) and (19021180095)

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
November 2022

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"Fitnesso: Fitness Score Predictor and AI Dietician"** in partial fulfillment of the requirements for the award of the Bachelors of Technology submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of July 2022, to December 2022, under the supervision of Dr. Arvind Kumar Professor, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Chirag Kaushik (19SCSE1010729) and

Sakshi Gaur (19SCSE1180106)

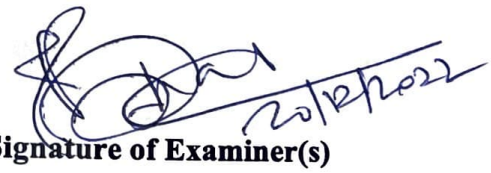
This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr. Arvind Kumar

Professor

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Chirag Kaushik and Sakshi Gaur has been held on 20/12/2022 and his/her work is recommended for the award of Bachelor of Technology in Computer Science and Technology.



Signature of Examiner(s)

Signature of Supervisor(s)

Signature of Program Chair

Signature of Dean

Date: December 2022

Place: Greater Noida

Abstract

Health is the real wealth of any individual, so it is important to be aware about it. Many times, it is observed that several people died because they came to know about their diseases at the last stage.

So, keeping a track of health status is important but regular visit to doctor is too much expensive. Hence, we developed a model that will brief you about your health status and gives you the proper measures that would be taken to help to improve and keep track of health progress.

The model works on different machine learning algorithms that gives you highly precise and accurate results. The designed model can predict the same results as given by highly expensive checkups which sometimes, are not affordable.

With high accuracy, the model is also time and resource efficient.

It's also evident many times, due to delay in results of tests, many patients lose their lives, but this model will give you quick results, helping you in saving your life from many diseases.

Along with this, you will be provided a diet plan and some physical exercises, by following which you can improve your health status, and keep a track of it.

Table of Contents

CHAPTER-1
Introduction

All around the globe, if you ask people about the health check, one thing that they will never miss to mention is that it is quite expensive and because of this very fact, some people are not even able to do so on a regular basis. Keeping this situation in mind, we have formulated the solution that we will build an app that will help people to get their fitness score along with their diet plans and exercises recommendation.

Now, it might look time consuming but it isn't. The user just need to login to the app and enter few inputs and it's done. It will provide you the score of your fitness and along with that it will also help you to be aware about chances of risk of some very common lifestyle diseases. Additionally, it will provide you the diet plan, and the set of exercises that you need to perform in order to improve your health.

For the regular users, it will provide various benefits that will just be a click away like the regular health updates, health related news, BMI calculator and much more.

For the execution of this, we will be using the concepts of several machine learning algorithms, web development, python etc.
Web development will help it to become usable so that even new users can use it easily, interactive and presentable.
Machine learning has proved to be one of the greatest tools in the prediction of crucial results in many fields like biomedical, analytical etc. So it will help us to predict more accurate results.
Python is quite handy and concise. It will ease the process of implementation of machine learning algorithms.

This project will be of long term use as health is the most important asset one can have. If he/she is well then it increases their productivity as well the efficiency. It will be of great use to all the people instead of just a part of population.

For updating, the user just need to let the app know about the changes and it will update the health status of the user in return. Hence, it is also easy to update.

CHAPTER-2
Literature Survey

On the back end, the model will be predicting the chances of various diseases that the user can have. As of now, we have included Heart Disease, Diabetes, Obesity, Hypertension and Breast Cancer. Related to these we have studied many of the research papers and following are the insights that we have attained-

- Papers based on Heart disease have revealed the fact that there can be several classes of attributes like age, gender, Thalach, CPT, FBS, Rest ECG etc. and the following models predicts the results quite accurately:

  SVM > Logistic Regression > Neural Network

- Research papers on Diabetes helps us to know that Type 2 diabetes is the only type of diabetes that we can predict as causes and prevention of Type 1 are unknown and 90% people are Type 2 patients only. Type 2 is caused by low insulin production or incorrect use of produced insulin. The set of attributes that are important here are BP, BMI, Skin, Age, Gender, etc. The models that can be used for this are:

  Logistic Regression > Decision Trees > Random Forest

- In the prediction of obesity, the attributes that can help us are Age, Gender, Height, Weight, Exercise Regularly, Smoke, Heart disease etc. After reading the papers on it, we came to know the following order of models as per the decreasing order of accuracies:

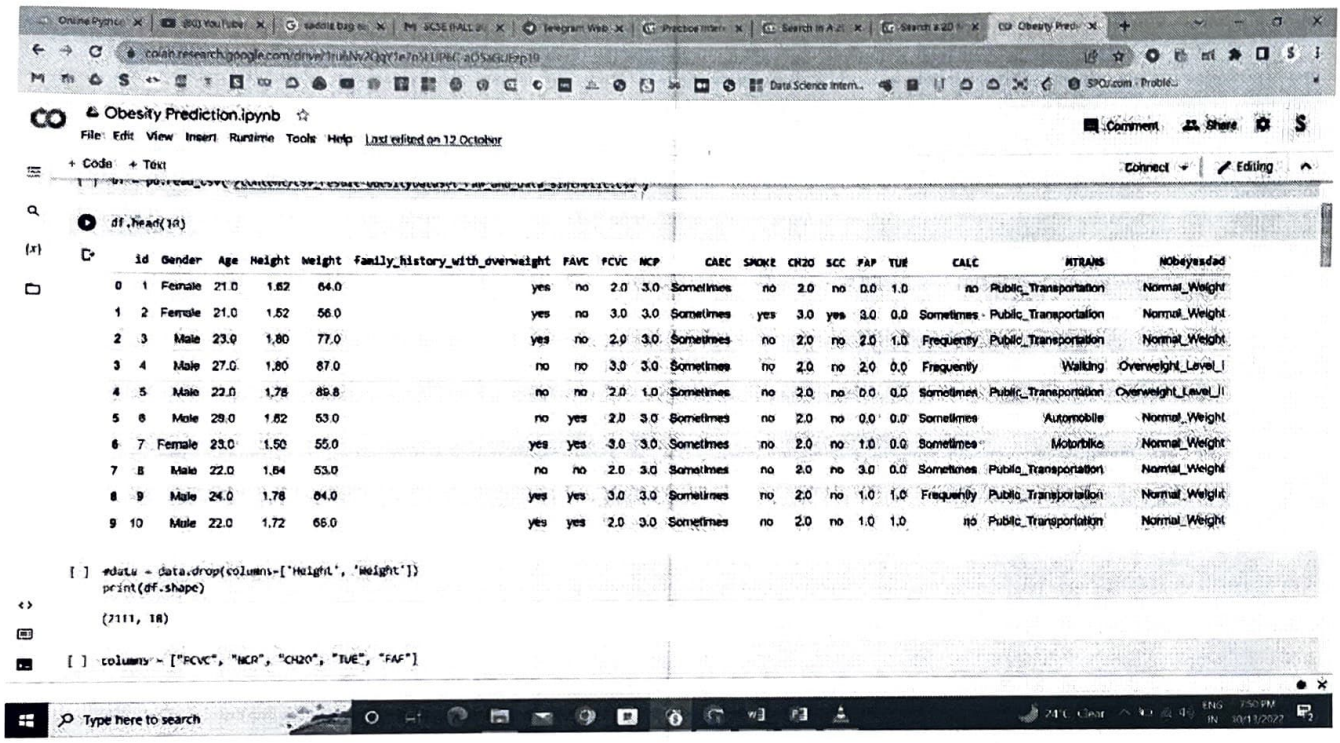
  Logistic Regression > Naïve Bayes > Decision Tree

- In case of hypertension, the direct connection of the disease is to the BP. The list of factors that can affect the prediction are BMI, BP, Cholesterol, Gender, Age and both types of BP (systolic and diastolic). The models that can be used for its prediction are:

  Neural Network > Logistic Regression > Random Forest

CHAPTER-3
Project Design

We have used several databases for various diseases and following are the snapshots of those for the reference in the order of Obesity, Heart Disease and Diabetes.



Obesity Prediction.ipynb

File Edit View Insert Runtime Tools Help Last edited on 12 October

+ Code + Text

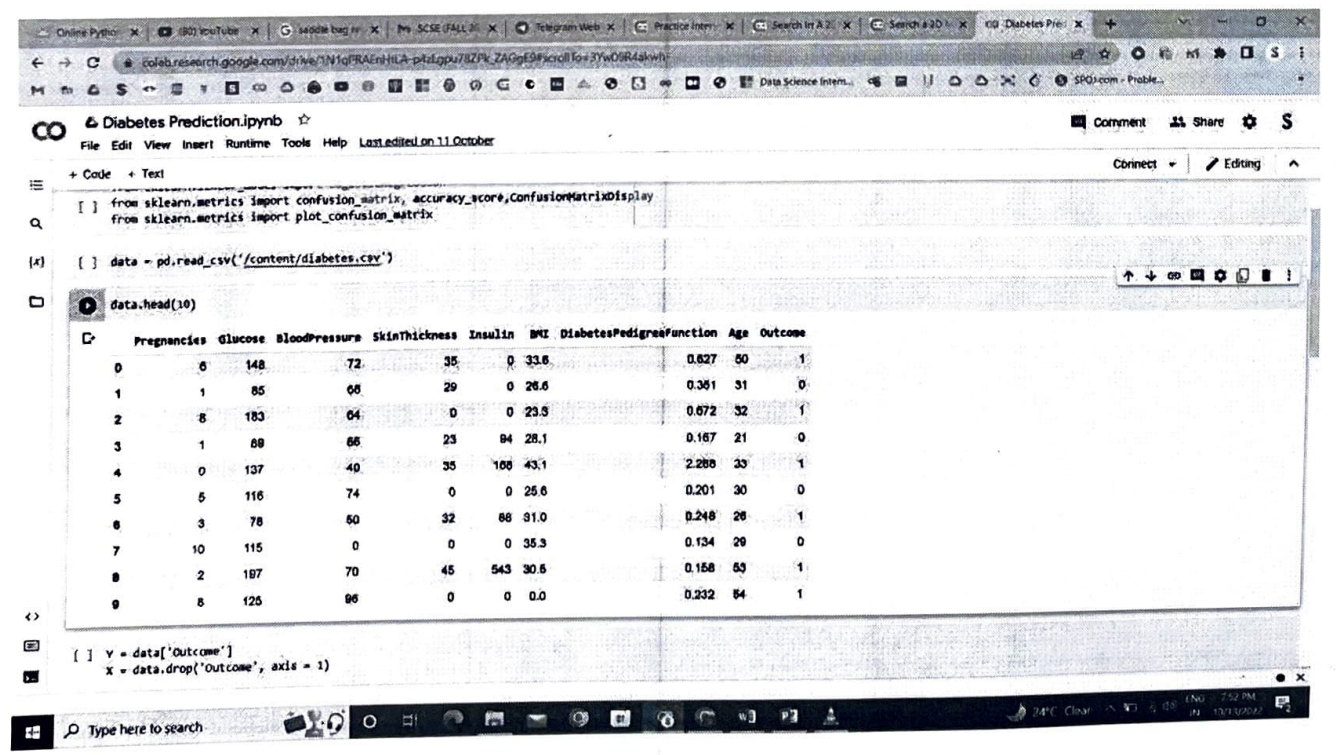
```
df.head(10)
```

	id	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
0	1	Female	21.0	1.62	64.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	0.0	1.0	no	Public_Transportation	Normal_Weight
1	2	Female	21.0	1.52	56.0	yes	no	3.0	3.0	Sometimes	yes	3.0	yes	3.0	0.0	Sometimes	Public_Transportation	Normal_Weight
2	3	Male	23.0	1.80	77.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	2.0	1.0	Frequently	Public_Transportation	Normal_Weight
3	4	Male	27.0	1.80	87.0	no	no	3.0	3.0	Sometimes	no	2.0	no	2.0	0.0	Frequently	Walking	Overweight_Level_I
4	5	Male	22.0	1.78	89.8	no	no	2.0	1.0	Sometimes	no	2.0	no	0.0	0.0	Sometimes	Public_Transportation	Overweight_Level_II
5	6	Male	29.0	1.62	53.0	no	yes	2.0	3.0	Sometimes	no	2.0	no	0.0	0.0	Sometimes	Automobile	Normal_Weight
6	7	Female	23.0	1.50	55.0	yes	yes	3.0	3.0	Sometimes	no	2.0	no	1.0	0.0	Sometimes	Motorbike	Normal_Weight
7	8	Male	22.0	1.64	53.0	no	no	2.0	3.0	Sometimes	no	2.0	no	3.0	0.0	Sometimes	Public_Transportation	Normal_Weight
8	9	Male	24.0	1.78	64.0	yes	yes	3.0	3.0	Sometimes	no	2.0	no	1.0	1.0	Frequently	Public_Transportation	Normal_Weight
9	10	Male	22.0	1.72	68.0	yes	yes	2.0	3.0	Sometimes	no	2.0	no	1.0	1.0	no	Public_Transportation	Normal_Weight

```
#data = data.drop(columns=['Height', 'Weight'])
print(df.shape)
```

```
(2111, 18)
```

```
columns = ["FCVC", "NCP", "CH2O", "TUE", "FAF"]
```



Diabetes Prediction.ipynb

File Edit View Insert Runtime Tools Help Last edited on 11 October

+ Code + Text

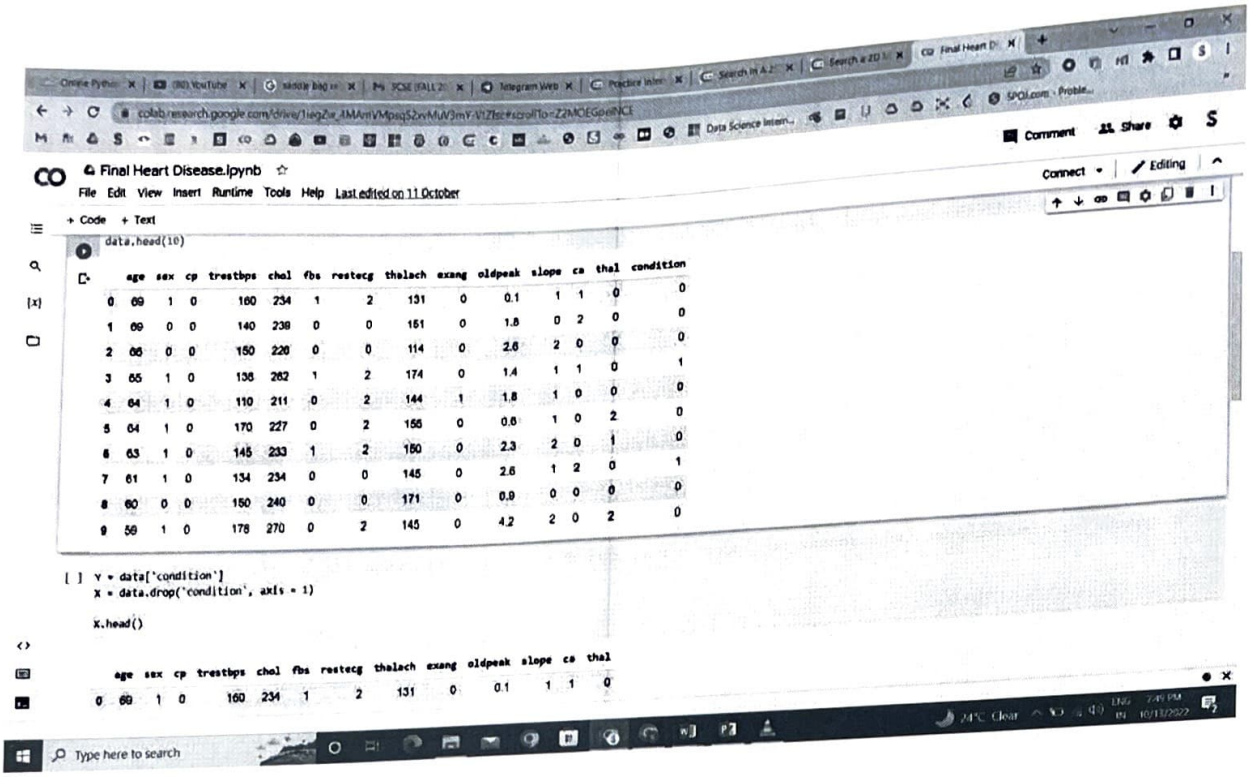
```
from sklearn.metrics import confusion_matrix, accuracy_score,ConfusionMatrixDisplay
from sklearn.metrics import plot_confusion_matrix
```

```
data = pd.read_csv('/content/diabetes.csv')
```

```
data.head(10)
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

```
Y = data['Outcome']
X = data.drop('Outcome', axis = 1)
```



Final Heart Disease.ipynb

File Edit View Insert Runtime Tools Help Last edited on 11 October

+ Code + Text

```
data.head(10)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	condition
0	69	1	0	160	234	1	2	131	0	0.1	1	1	0	0
1	69	0	0	140	238	0	0	151	0	1.8	0	2	0	0
2	66	0	0	150	226	0	0	114	0	2.6	2	0	0	0
3	65	1	0	138	282	1	2	174	0	1.4	1	1	0	1
4	64	1	0	110	211	0	2	144	1	1.8	1	0	0	0
5	64	1	0	170	227	0	2	155	0	0.6	1	0	2	0
6	63	1	0	145	233	1	2	150	0	2.3	2	0	1	0
7	61	1	0	134	234	0	0	145	0	2.6	1	2	0	1
8	60	0	0	150	240	0	0	171	0	0.9	0	0	0	0
9	59	1	0	178	270	0	2	145	0	4.2	2	0	2	0

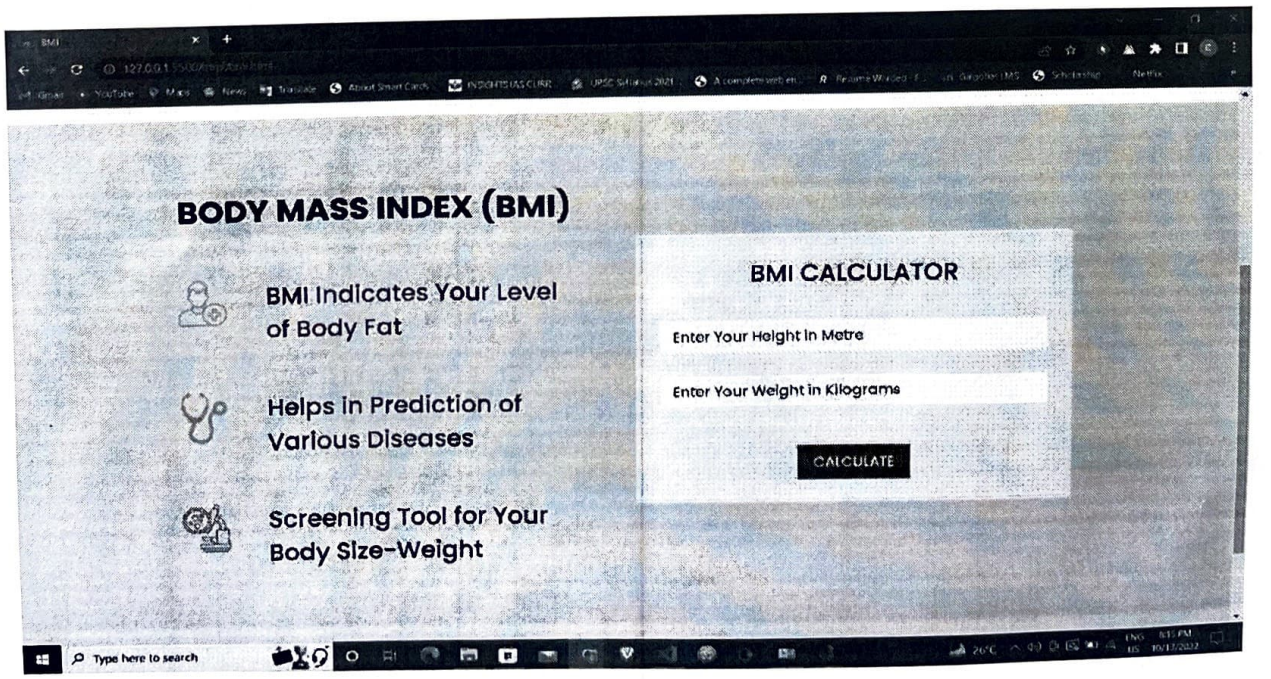
```
Y = data['condition']
X = data.drop('condition', axis = 1)

X.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	69	1	0	160	234	1	2	131	0	0.1	1	1	0

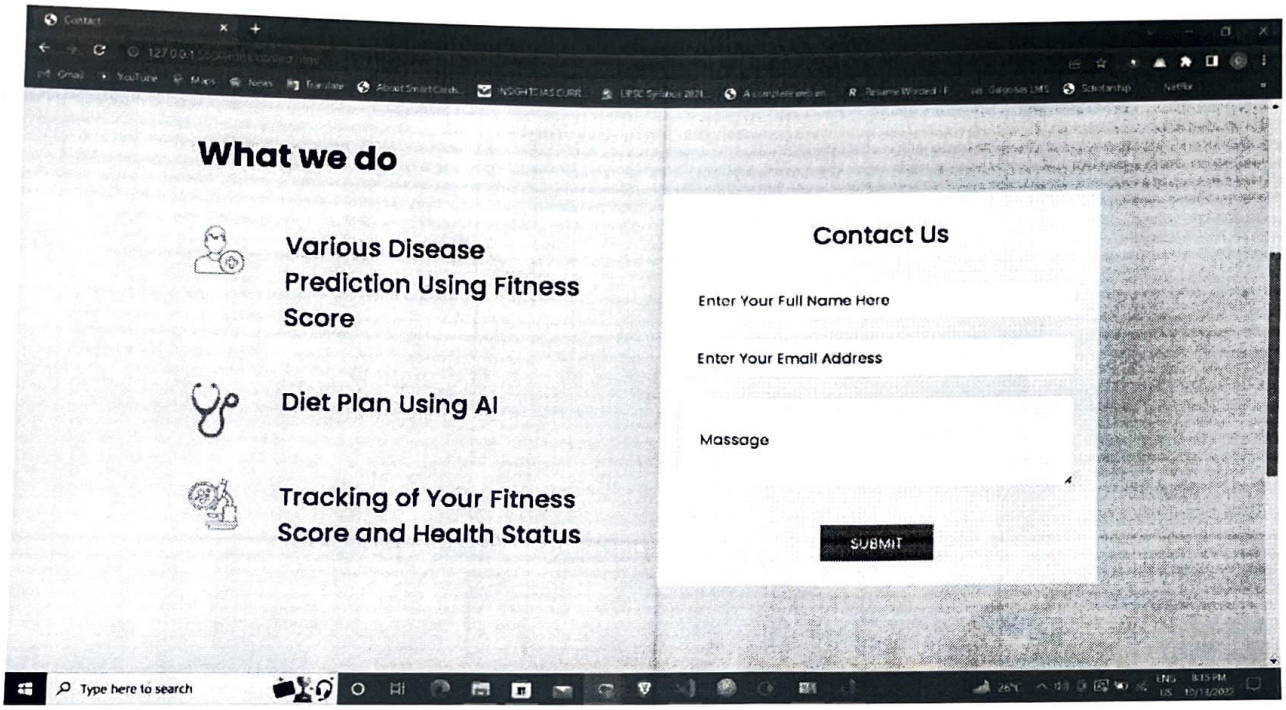
CHAPTER-4
Module Description

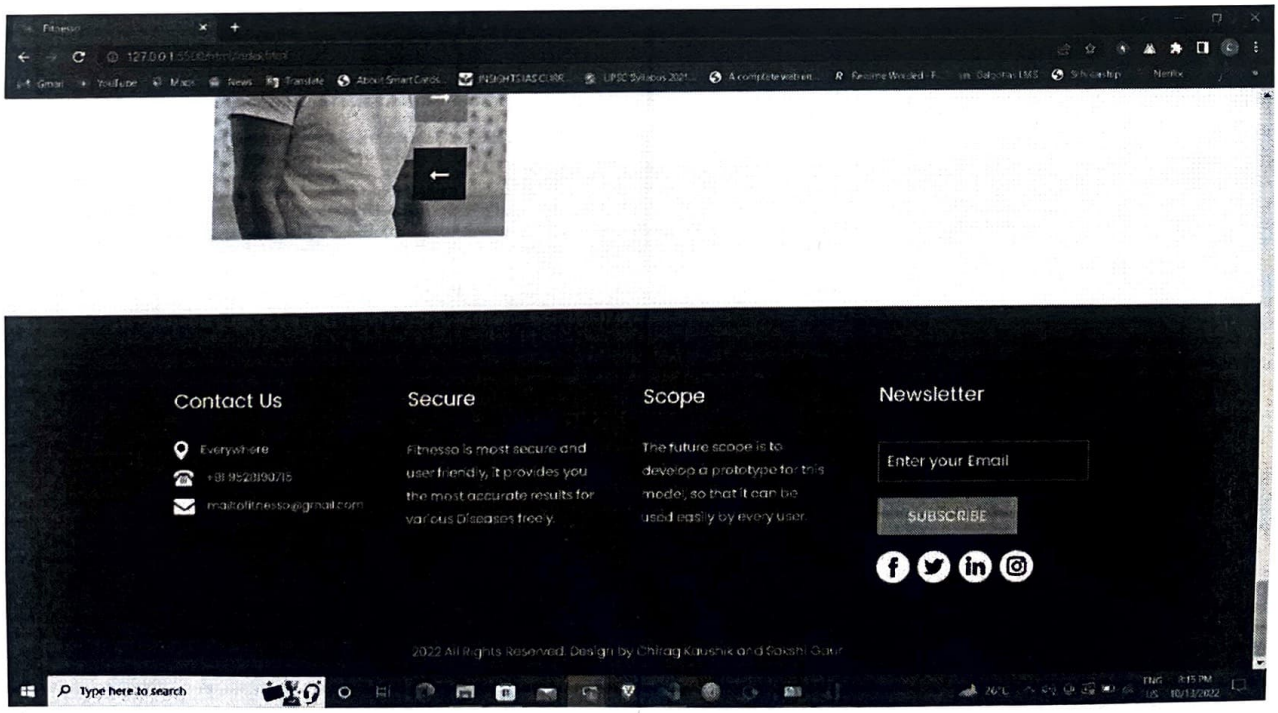
This is the API that we have built for the app. It will have options as demonstrated in the screenshot like Home, FI Score, BMI, Medicine, News, Client, and Contact Us.

The following image shows the interface that helps to calculate the BMI for user with the recommendation of whether the user is normal, underweight or overweight and their respective recommendations.

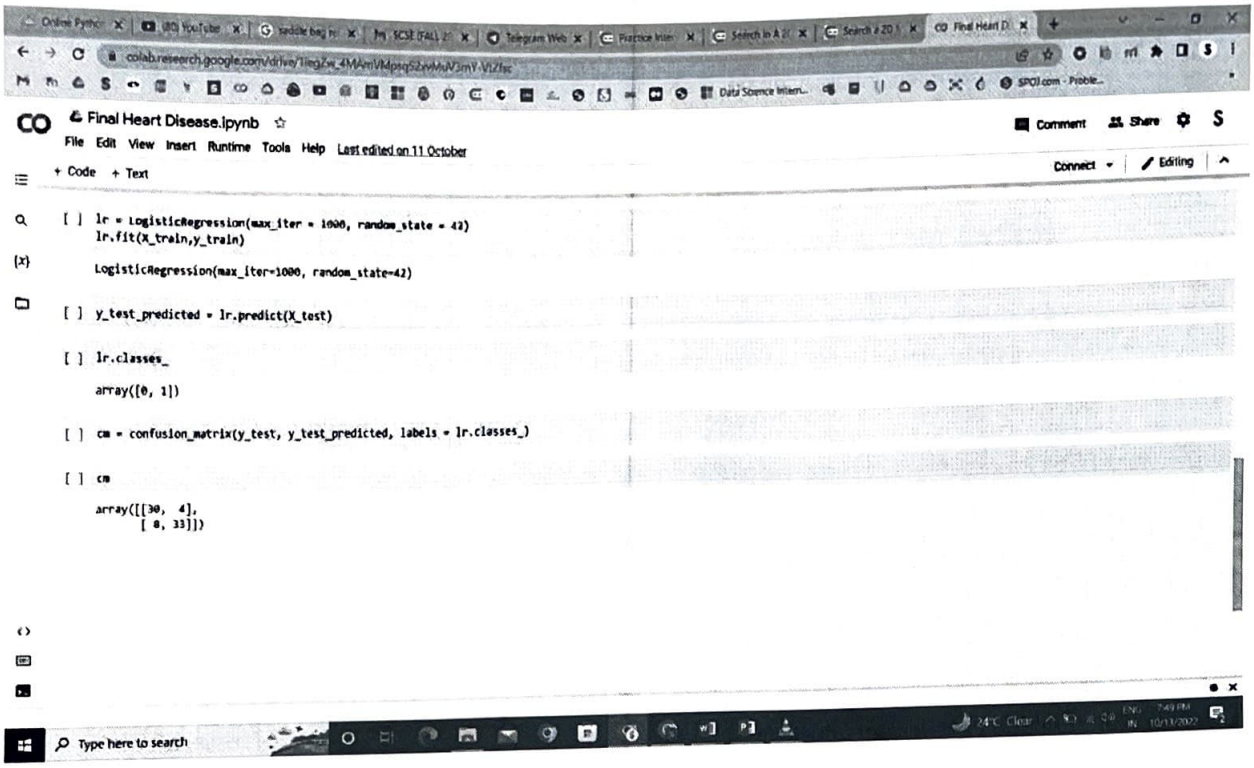


This interface helps us to connect with the clients so that they can provide feedback to us and we will make it better for future use.

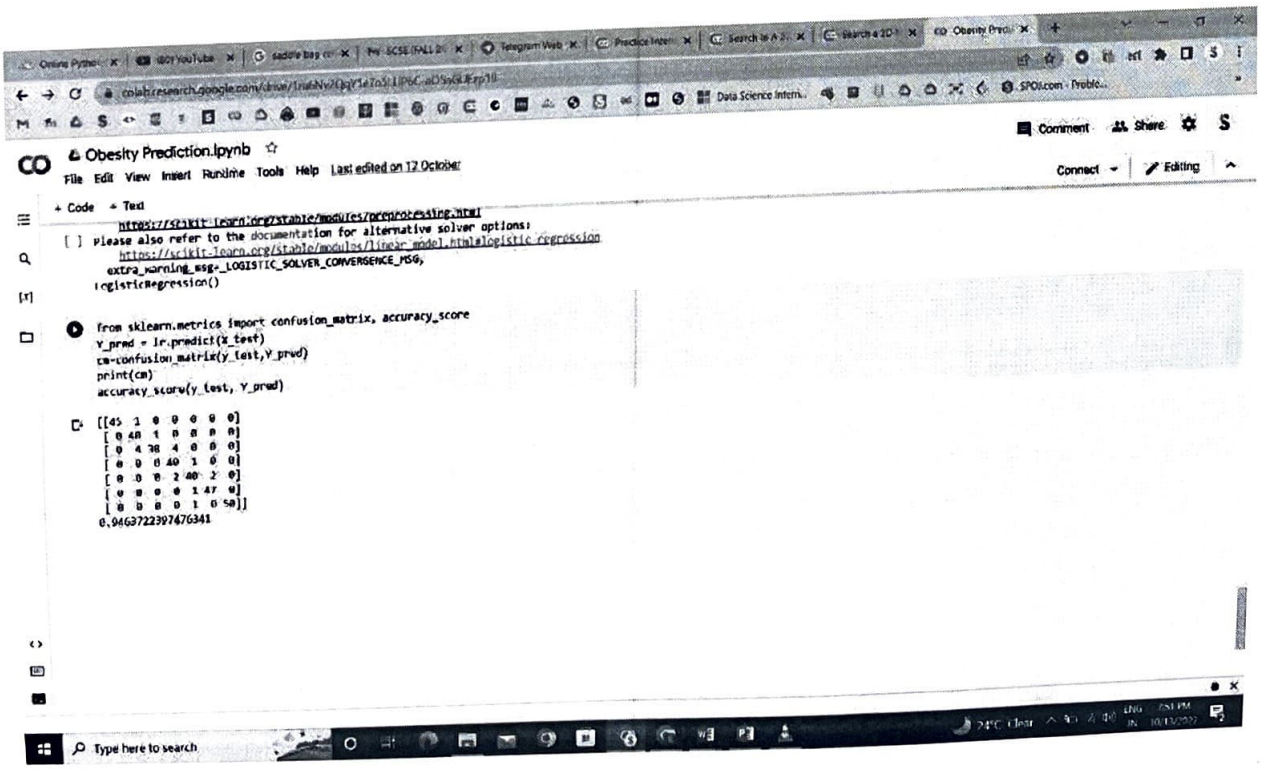




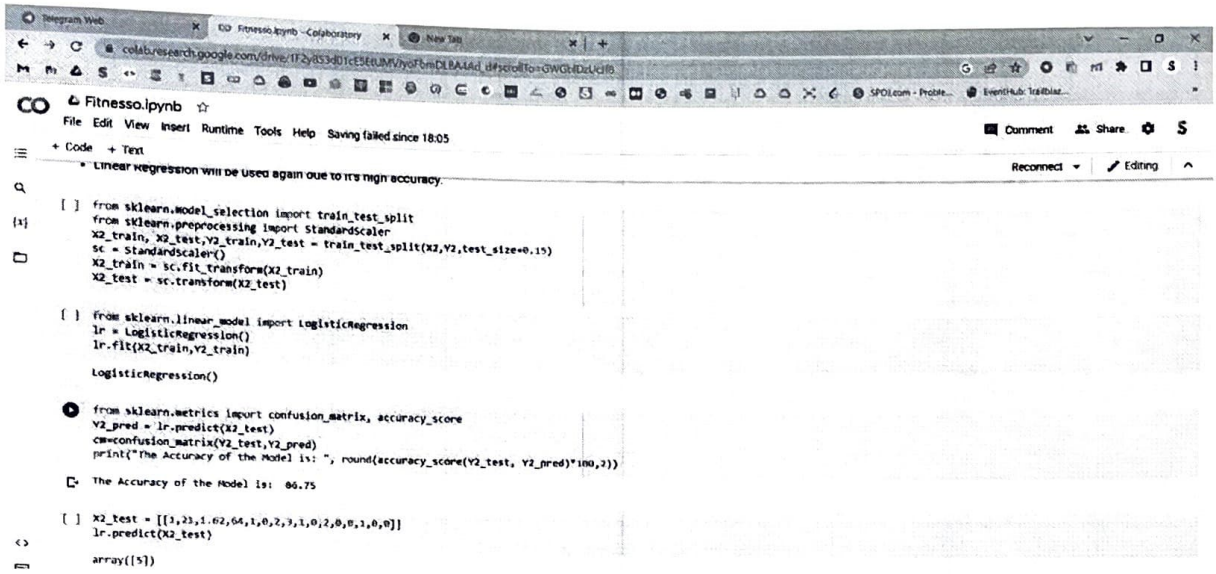
This is the model that helps to determine whether the user is at risk of having a heart disease or not. The screenshot has also displayed the results of the model in the form of confusion matrix. The classification model used here is logistic regression.



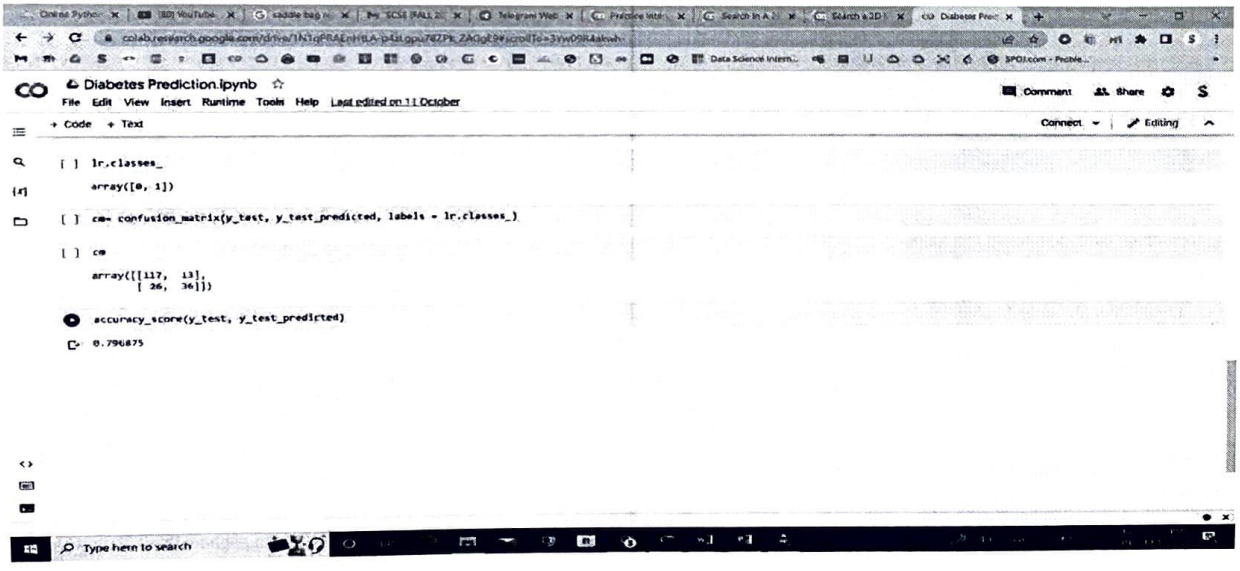
The dataset available for the disease obesity was impure. We have preprocessed the data, done the label encoding for a number of attributes and filled the missing values with the statistical ones like mean, median and mode. After that, the model is trained on logistic regression which gives us the final accuracy of 94.6%.



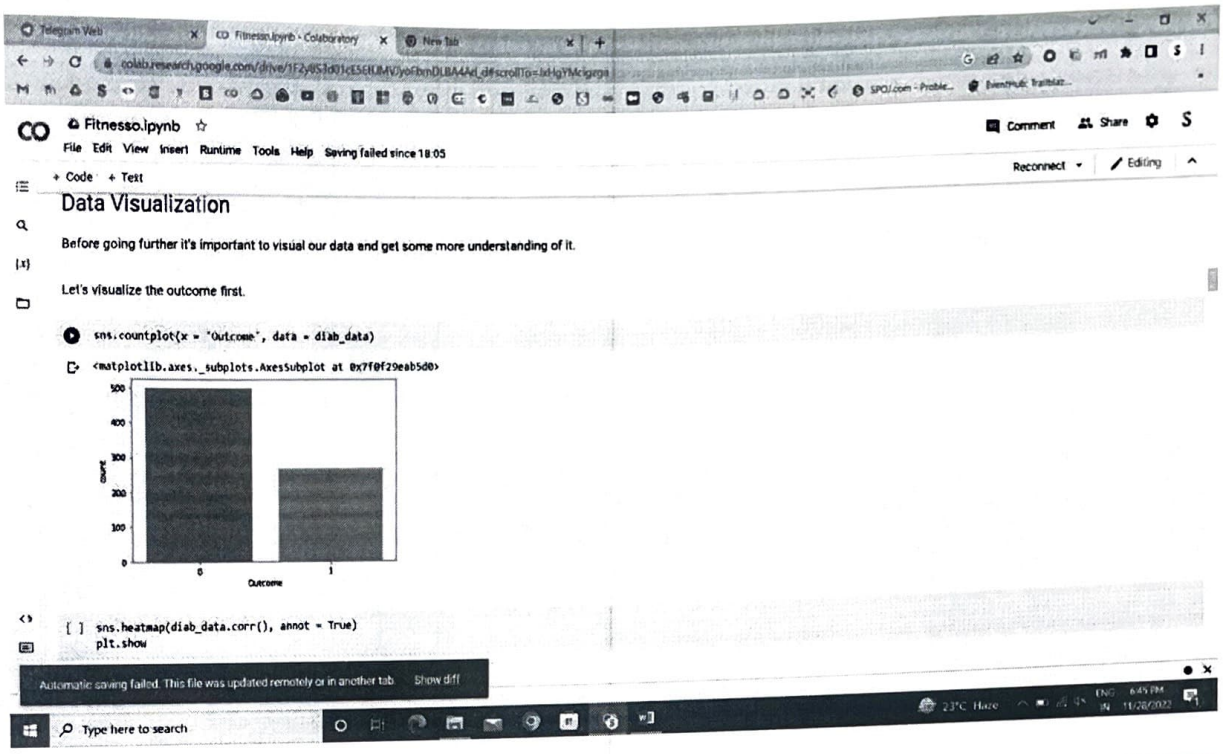
The same dataset is trained on linear regression because the previous model is over fit and now it is best fit at the cost of sacrificing test data accuracy a bit.

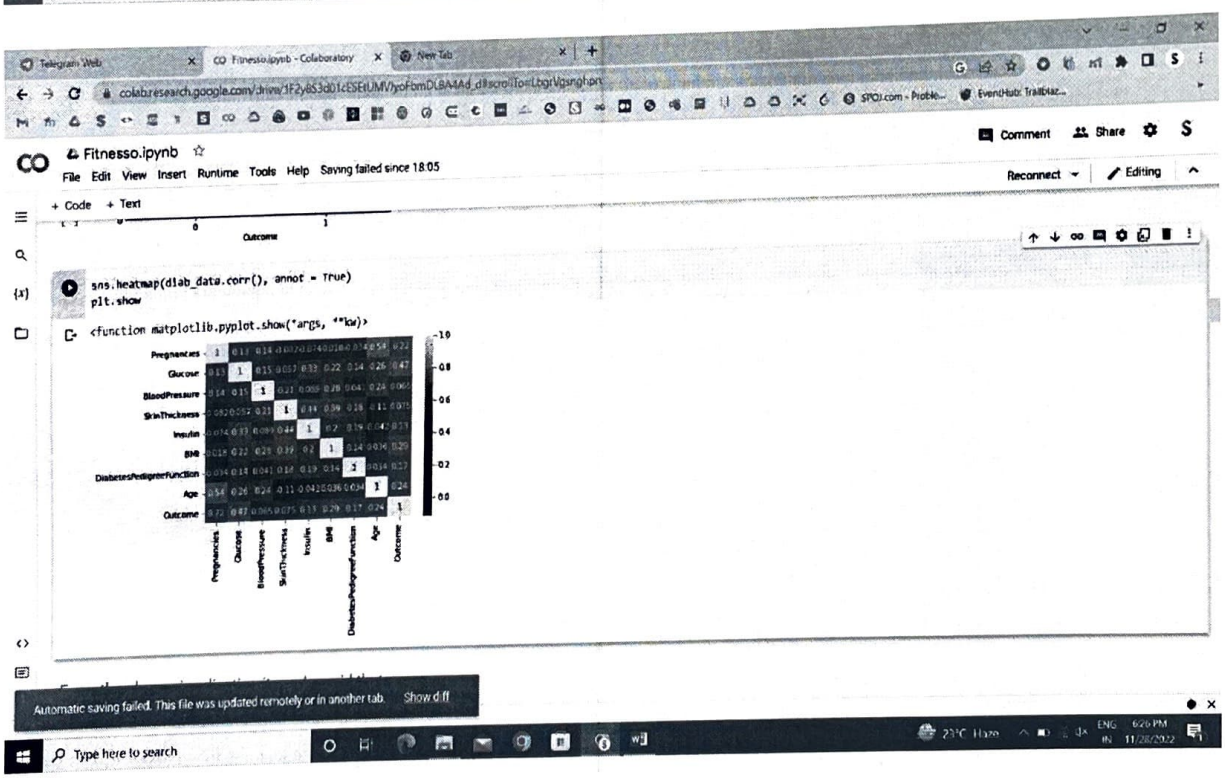


The prediction model of diabetes is trained on the clean dataset. And the logistic regression classification model is used to classify whether a person is diabetic or not giving an accuracy of 79.6%.



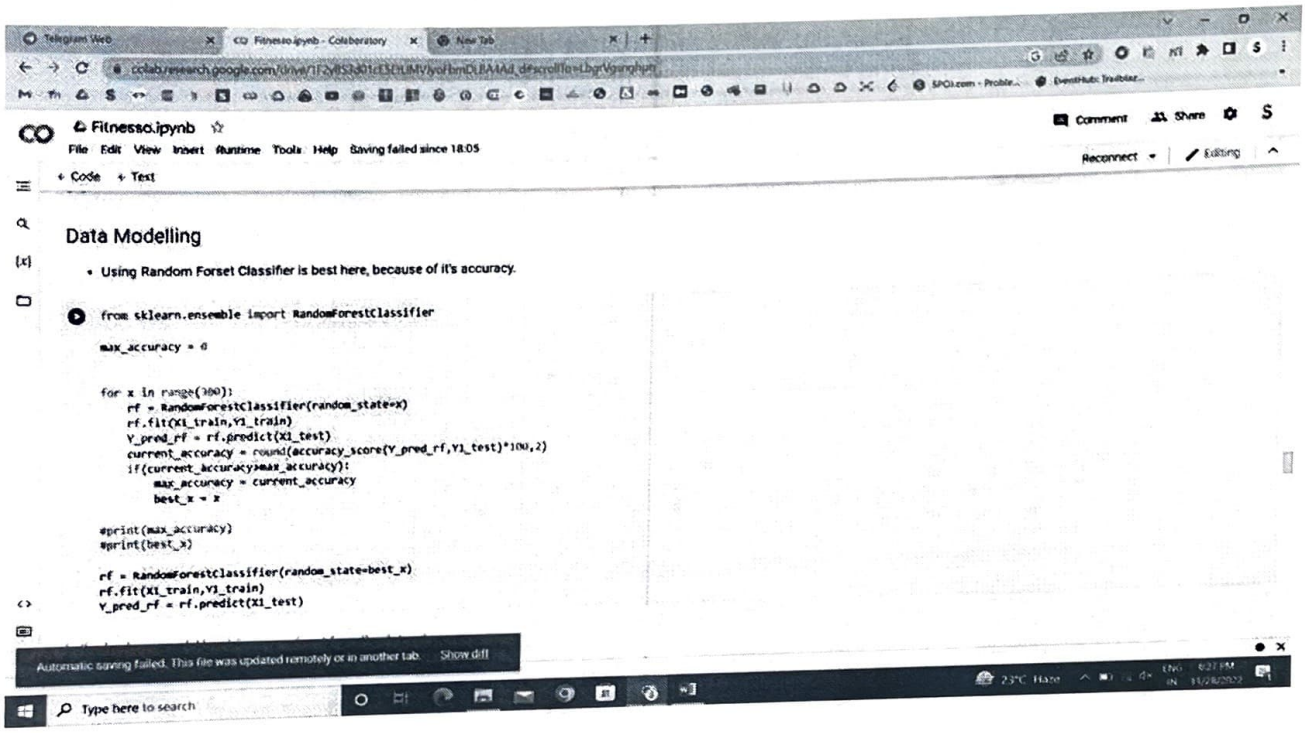
The prediction model for diabetes has better accuracy over random forest after reading the heat map of it. The screenshot of the count plot, heat map and the implementation is shown below for the reference.





From the above visualization it can be said that:
Count plot shows us the dataset is imbalanced between the patients with diabetes and patients with no diabetes.

Correlation heat map shows us that the "Outcome" is highly correlated with Glucose, BMI, Insulin and Age. So, these features can be used as input to predict the outcome.
Less features will not provide accurate results so here we'll be taking all the features to predict the target variables.



REFERENCES:

[1] Romeshwar Sookrah, Jaysree Devee Dhowtal and Soulakshmee Devi Nagowah, "A DASH Diet Recommendation System for Hypertensive Patients Using Machine Learning",2019 7th International Conference on Information and Communication Technology.

[2] Gergely Kov´asznai, "Developing an expert system for diet recommendation", 2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics.

[3] Wahidah Husain, Lee Jing Wei, Sooi Li Cheng and Nasriah Zakari,"Application of data mining techniques in a personalized diet recommendation system for cancer patients", 2011 IEEE Colloquium amd Humanities,Science and Engineering.

[4] Ashivini Kale and Nisha Auti, " Automated Menu Planning Algorithm for Children: Food Recommendation by Dietary Management System using ID3 for Indian Food Database", Procedia Computer Science,Volume 50, 2015, pp 197-202.

[5]A. Aamodt, E. Plaza, "Case-based Reasoning, Foundation Issues,Methodological Variations and System Approaches", AICom–Artificial Intelligence Communication, IOS Press, Vol.7, No. 1, pp 39-59, 1994.

[6] Jen-Hao Hsiao1 and Henry Chang2,"SmartDiet: A Personal Diet Consultant for Healthy Meal Planning ",IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS),Oct. 2010.

[7] Jong-Hun Kim, Jung-Hyun Lee, Jee-Song Park, Young-Ho Lee, Kee-Wook Rim, "Design of Diet Recommendation System for Healthcare Service Based on User Information", 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, 2009.

[8] Sakshi Singh, Sanjay Kumar Dubey, "Recommendation of Diet to a Patient using AHP and Fuzzy Approach", 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019.

[9] Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol.2011;3:67.

[10] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machinelearning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944.

[11] Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol.2018;7(2.8):684–7.

[12] Pouriyeh S, Vahid S, Sannino G, De Pietro G, Arabnia H, Gutierrez J. A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In: 2017 IEEE symposium on computers and communications (ISCC). IEEE. p. 204–207.

[13] Bouali H, Akaichi J. Comparative study of different classification techniques: heart disease use case. In: 2014 13th international conference on machine learning and applications. IEEE. p. 482–86.