**FIT & HEALTHY**

A PROJECT REPORT

***Submitted by***

# MRIDUL DESAI (20MIM10010)

# KRISHNA KUMAR GUPTA (20MIM10022) SHAGUN SRIVASTAVA (20MIM10061) SNEHA PRASAD (20MIM10069)

*in partial fulfillment for the award of the degree of*

***DSN2098 Project Exhibition-I***

# MASTER OF TECHNOLOGY

*in*

**COMPUTER SCIENCE AND ENGINEERING with AIML**



**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

# VIT BHOPAL UNIVERSITY

# KOTHRIKALAN, SEHORE

# MADHYA PRADESH - 466114

29 April, 2022

**VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE MADHYA PRADESH – 466114**

# BONAFIDE CERTIFICATE

Certified that this project report titled **“FIT & HEALTHY”** is the Bonafede work of **“MRIDUL DESAI (20MIM10010), KRISHNA KUMAR GUPTA (20MIM10022), SHAGUN SRIVASTAVA (20MIM10061), SNEHA PRASAD (20MIM10069)”** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

|  |  |
| --- | --- |
| **PROGRAM CHAIR**  Dr Mayuri AVR  Senior Assistant Professor  Head of the Department  Int. M. Tech. (AIML)  School of Computer Science and Engineering VIT BHOPAL UNIVERSITY | **PROJECT GUIDE**    Dr. Pazhaniraja N  Teaching Fellow  Department of Computer Science  School of Computer Science and Engineering  VIT BHOPAL UNIVERSITY |

The Project Exhibition II Examination is held on **29th APRIL 2021**

# ACKNOWLEDGEMENT

First and foremost, I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr. Mayuri A V R, Head of the Department, Integrated M. Tech-AI, School of Computing Science and Engineering for much of his valuable support encouragement in carrying out this work.

I would like to thank my internal guided Mr. Pazhaniraja N, Teaching Fellow, Department of Computer Science, School of Computer Science and Engineering for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Computer Science, who extended directly or indirectly all support.

Last, but not least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

# LIST OF FIGURES AND GRAPHS

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Figure No. | Title | Page No. | | 1 | User Interface design | 16 | | 2 | Flow-diagram | 18 | |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**ABSTRACT**

k-Nearest Neighbor (kNN) algorithm is an effortless but productive machine learning algorithm. It is effective for classification as well as regression. However, it is more widely used for classification prediction. kNN groups the data into coherent clusters or subsets and classifies the newly inputted data based on its similarity with previously trained data. The input is assigned to the class with which it shares the nearest neighbors. Though kNN is effective, it has many weaknesses. This project uses the kNN method and its modified versions available in previously done researches. These variants remove the weaknesses of kNN and provide a more efficient method.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | Related Work Investigation | 8 |
| 1 | **CHAPTER-1:**  **PROJECT DESCRIPTION AND OUTLINE**  1.1 Motivation for the work  1.2 Introduction to project  1.3 Problem Statement  1.4 Objective of the work | 9 |
| 2 | **CHAPTER-2:**  **RELATED WORK INVESTIGATION**  2.1 Introduction  2.2 Core area of the project  2.3 Existing Research works  2.4 Pros and cons of the stated Research Works  2.5 Observations from investigation  2.6 Summary | 11 |
| 3 | **CHAPTER-3:**  **REQUIREMENT ARTIFACTS**  3.1 Hardware requirements  3.2 Software requirements  3.3 Specific Project requirements | 14 |
| 4 | **CHAPTER-4:**  **DESIGN METHODOLOGY AND ITS NOVELTY**  4.1 Methodology and goal  4.2 Functional modules design and analysis  4.3 User Interface designs | 16 |
| 5 | **CHAPTER-5:**  **TECHNICAL IMPLEMENTATION & ANALYSIS**  5.1 Working Layout 1  5.2 Working Layout 2 | 18 |
| 6 | **CHAPTER-6:**  **PROJECT OUTCOME AND APPLICABILITY**  6.1Test and Validation  6.2 Key implementations outline of the System  6.3 Significant project outcomes  6.4 Project applicability on Real-world applications | 23 |
| 7 | **CHAPTER-7:**  **CONCLUSIONS AND RECOMMENDATION**  7.1Outline  7.2 Future Enhancements | 24 |
| 8 | **References** | 25 |

# 

## RELATED WORK INVESTIGATION

The emergence of Artificial Intelligence (AI) transformed the world in all the fields. Machine learning (ML), a subset of AI helps the human to find solutions for highly complex problems and also plays a vital role in making human life sophisticated. The application areas of ML include business applications, intelligent robots, autonomous vehicle (AV), healthcare, climate modelling, image processing, natural language processing (NLP), and gaming. The learning of ML mimics human intelligence, and it is implemented based on trial-and-error method. The instructions to the algorithm were given mainly using control statements such as conditional if, many prediction-based algorithms are available in ML. The ML techniques are used for classification and prediction in various fields like disease prediction, stock market, weather forecasting, and business. In the medical field also, many ML algorithms are used for disease prediction like coronary artery disease, predicting cardiovascular disease, and prediction of breast cancer. Several researches are also done for COVID-19 confirmed case live forecasting and for predicting the COVID-19 outbreak. These works will aid the higher authorities of the country for taking decisions to handle the situation by foreseeing. At first, the COVID-19 was misinterpreted as pneumonia. But the failure of multi-organs and high mortality rates made it a pandemic in the whole world. Classification techniques are broadly categorized into semi-supervised, supervised and unsupervised. Supervised learning takes information about the classes and learns based on that information. Based on this knowledge, this technique can predict the classes for new data. In unsupervised learning, the information about the classes is unknown. The clustering of similar data is done by identifying the similarity among themselves. Semi supervised techniques know some information about data, and the classification is done based on it. Logistic Regression is used for relationship analysis between various dependent variables.

**CHAPTER-1:**

**PROJECT DESCRIPTION AND OUTLINE**

## 

## Motivation for the work

We have got the motivation of doing this project after looking at the state of hospitals and clinics at the time of covid or even at general days. The amount of time a person has to wait even for a simple test I unbelievable. So, we as a group thought that why not to make a project by which one can do the simple test at their home itself, henceforth we made this project.

## 1.2 Introduction to Project

In our project one can do the very simple medical tests such as checking their BP or Sugar Level just by sitting at their very own home. This does reduce their precious time and also saves the clinics from getting crowded.

## 1.3 Problem Statement

1. **To minimize the time consumption**

  Saving the precious time of doctors and other supporting staff.

**2. Reducing the hospital visits**

Visiting hospitals is quite risky during pandemic hence our program will help in reducing it to a certain level.

**3. Self-help is the best help**

   Less knowledge about medicines/home remedies problem, we’ve got you covered.

**1.4 Objective of Work**

Our objective of Fit&Healthy app is to improve medical assistance and patients’ recovery process.

**CHAPTER-2**

**RELATED WORK INVESTIGATION**

## 2.1 Introduction

## Now a days visiting doctor's clinic is not that safe therefore people are preferring to stay at home and check their ailment through self-testing kits such as " my lab covi-self kit " for covid 19 , " glucometer " for blood sugar level , " oximeter " for oxygen levels , " sphygmomanometer " for measuring blood pressure level , etc. Our program will help them to know whether the values/indications given by the above kits/instruments are below/above normal or normal along with this we'll provide some home remedies to cure the ailment such as caffeine intake can help in increasing blood pressure level hence bringing thus bringing it to a normal level.

**2.2 Core area of the project**

Health & Science: This field of study aims to develop knowledge, interventions and technology for use in healthcare to improve the treatment of patients. In our project we aim to use technology to make people’s life easy save the precious time of our life saver “Doctors”.

**2.3 Existing Research Works**

A disease predictor can be called a virtual doctor, which can predict the disease of any patient without any human error. Also, in conditions like COVID-19 and EBOLA, a disease predictor can be a blessing as it can identify a human’s disease without any physical contact.

Several models were initiated by using various machine learning (ML) algorithms that collected raw data and then bifurcated it according to gender, age group, and symptoms. The data-set was then processed in several ML models like Fine, Medium and Coarse Decision trees, Gaussian Naïve Bayes, Kernel Naïve Bayes, Fine, Medium and Coarse KNN, Weighted KNN, Subspace KNN, and RUS Boosted trees. According to ML models, the accuracy varied. While processing the data, the input parameters data-set was supplied to every model, and the disease was received as an output.

**2.4 Pros and cons of the stated Research work**

**Advantages**

* Reducing hospital visits:
* Visiting hospitals is quite risky during pandemic hence our program helps in reducing it to a certain level
* Minimizing the time consumption
* Doctor had a very tight packed schedule during the pandemic so saving their precious time such that they can invest their time in more important causes.

**Disadvantages**

* You cannot receive and expert advice on your condition /results updated.
* The suggestions and recommendations ​given by the program may or may not work ​for a particular individual.​
* There could be an error in the imputed data, as ​Human errors are possible and thus the result ​displayed would not be very accurate.

    ​

**2.5 Observations from investigation**

* In our project we have done the Analysing, training and testing the data /dataset of the data using knn implementation. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
* We have used tkinter for creating or making the interface for the user in order to access out program.
* We also have connected the interface with the program in order to produce desired outputs.
* Suppose there are two categories, i.e., Category A and Category B, and we have a new data point X1, so this data point will lie in which of these categories. To solve this type of problem, we need a KNN algorithm. With the help of KNN, we can easily identify the category or class of a particular dataset.

**2.6 Summary**

We have used kNN algorithm to classify and predict the accurate results of our data variables which includes classification of blood pressure, blood sugar levels, oxygen level , etc accurately by using support vector machine for categorizing and random forest method for prediction

* Blood Pressure - We take in the input from the user and classify it into low, high and normal range.
* Oxygen - We take in a single value from the user and classify it into low or normal
* Glucose level - We take in input from the user and demonstrate the output as higher than normal, lower than normal or normal.

Our program is 98% accurate in predicting the results.

**CHAPTER-3**

**REQUIREMENT ARTIFACTS**

## 3.1 Hardware Requirements

**INTEL/RYZEN**

* Intel i5 8th GEN or Ryzen 5 3600 chip or latest

**RAM**

* 4 GB internal RAM or above

**GRAPHIC CARD**

* 2 GB internal Graphic card or above

## 3.2 Software Requirements

The OS helps you to communicate with the computer without knowing how to speak the computer's language. It is not possible for the user to use any computer or mobile device without having an operating system. Majority of home users use a Windows based machine.

* Windows 7 or later
* Google chrome

## 3.3 Specific Project Requirements

* Data Requirements- 1 MBPS Speed
* Python Library
* Performance and security Requirements
* Windows Firewall
* Network Firewall

**CHAPTER-4**

**DESIGN METHODOLOGY AND ITS NOVELTY**

## 4.1 Methodology and goal

**Methodology:**

So, the Methodology used is to find the input of the data given my user using the Tkinter program. We use many testings, predicting and classification algorithms like SVM (support vector machine) and random forest algorithm to classify and efficiently predict the data using Tkinter, then we connected the program to a simple GUI made by the help of Tkinter.

**Goal:**

Our goal is to efficiently predict the results of the given input and classify the data to high, low or normal levels. So that the Users can efficiency manage their blood pressure or sugar levels conveniently at home itself without having to consult a doctor. we aim on providing maximum efficiency.

### 4.2 Functional Modules and analysis

The different modules used in program is pandas, Matplotlib, NumPy and seaborn pandas -pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

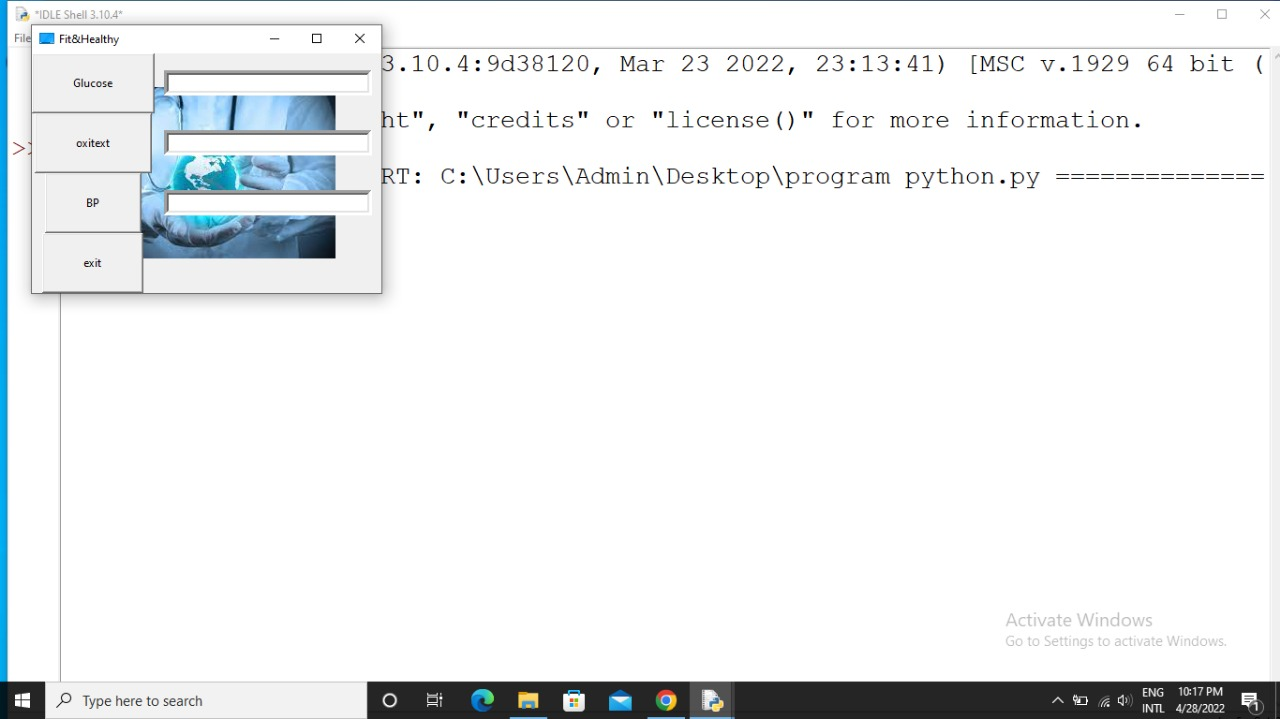
Matplotlib-Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, python, Qt, or GTK.

NumPy-NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

seaborn-Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas’ data structures. Seaborn helps you explore and understand data.

**4.3 User interface design:**

we have used the GUI based interface made using Tkinter which is very easy and user-friendly design so that anyone can access and utilize our program conveniently. we have the options to choose from and the users can simply put the calculated value in the box and run the program and get the desired outputs.



**CHAPTER-5**

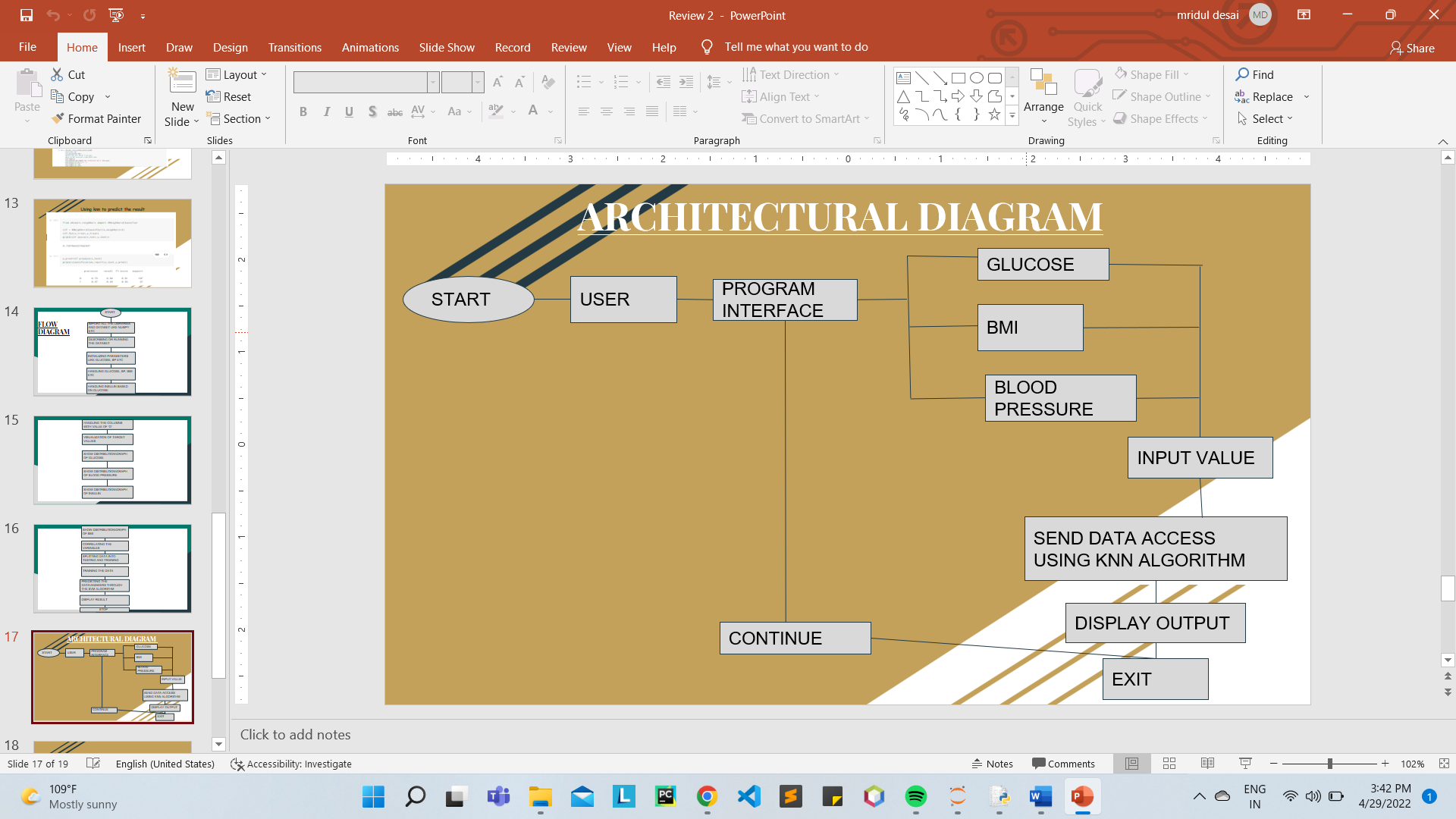
**TECHNICAL IMPLEMENTATION & ANALYSIS**

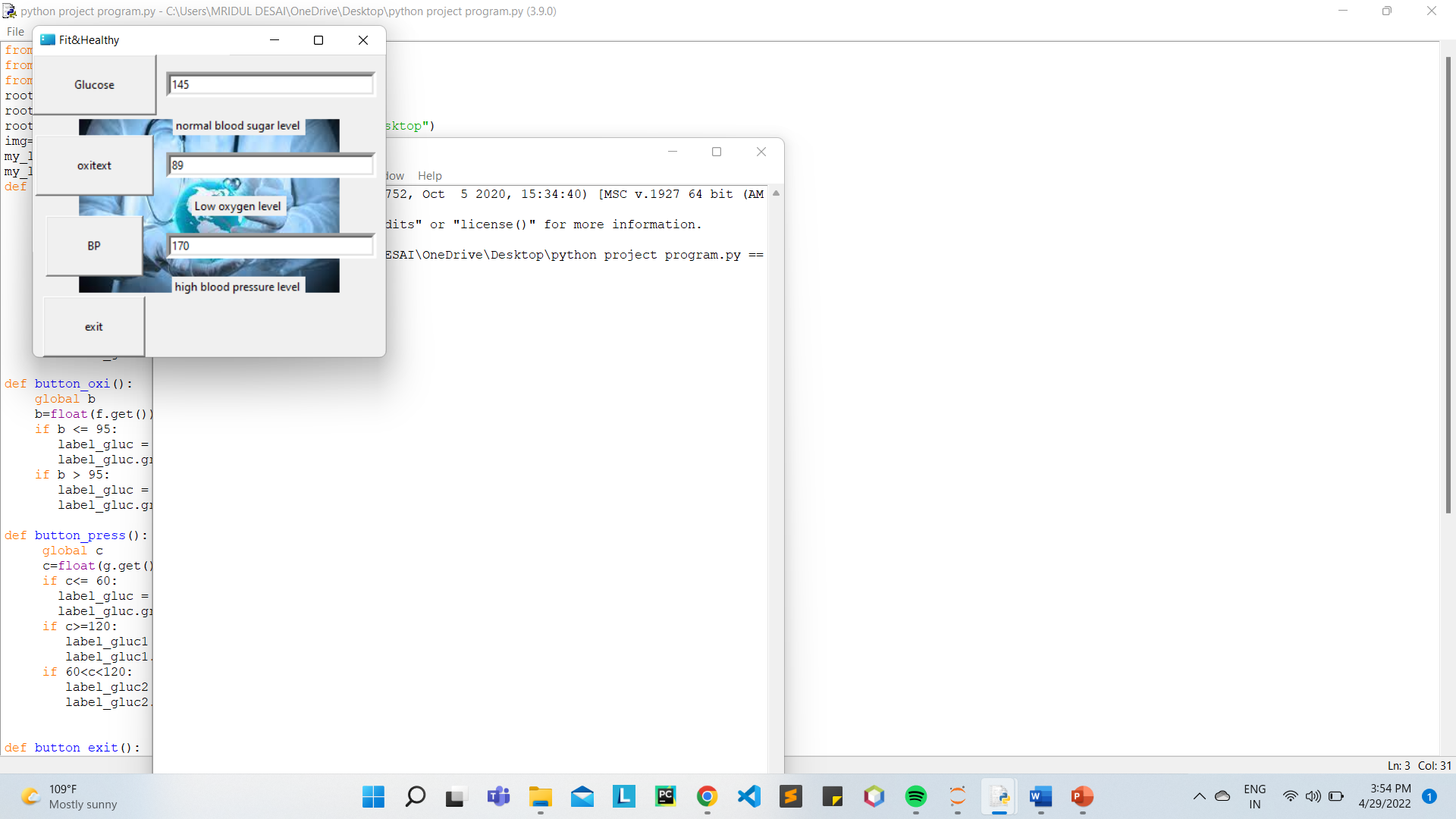
# 5.1 Working layouts of forms

* As we can see in the above figure is the working outline of our GUI.
* First the user inters into our program or the GUI.

Then is the home page having many options to choose from in it.

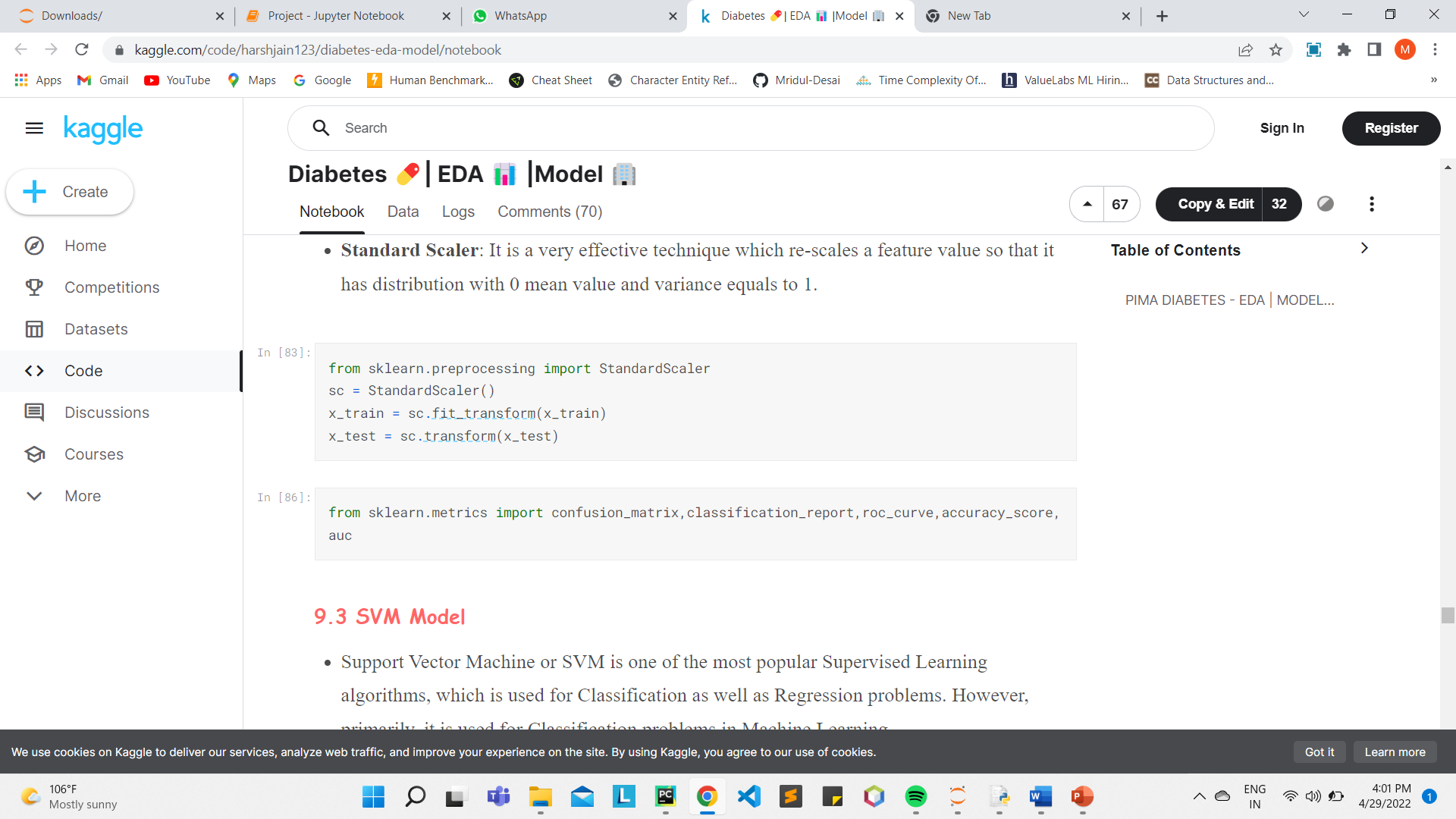
* Once the User gets into the homepage, then they can get to know about any of the medical test they want to check, they just have to enter the reading taken.
* Depending on the given input the Knn program runs and the suitable output is predicted and displayed.



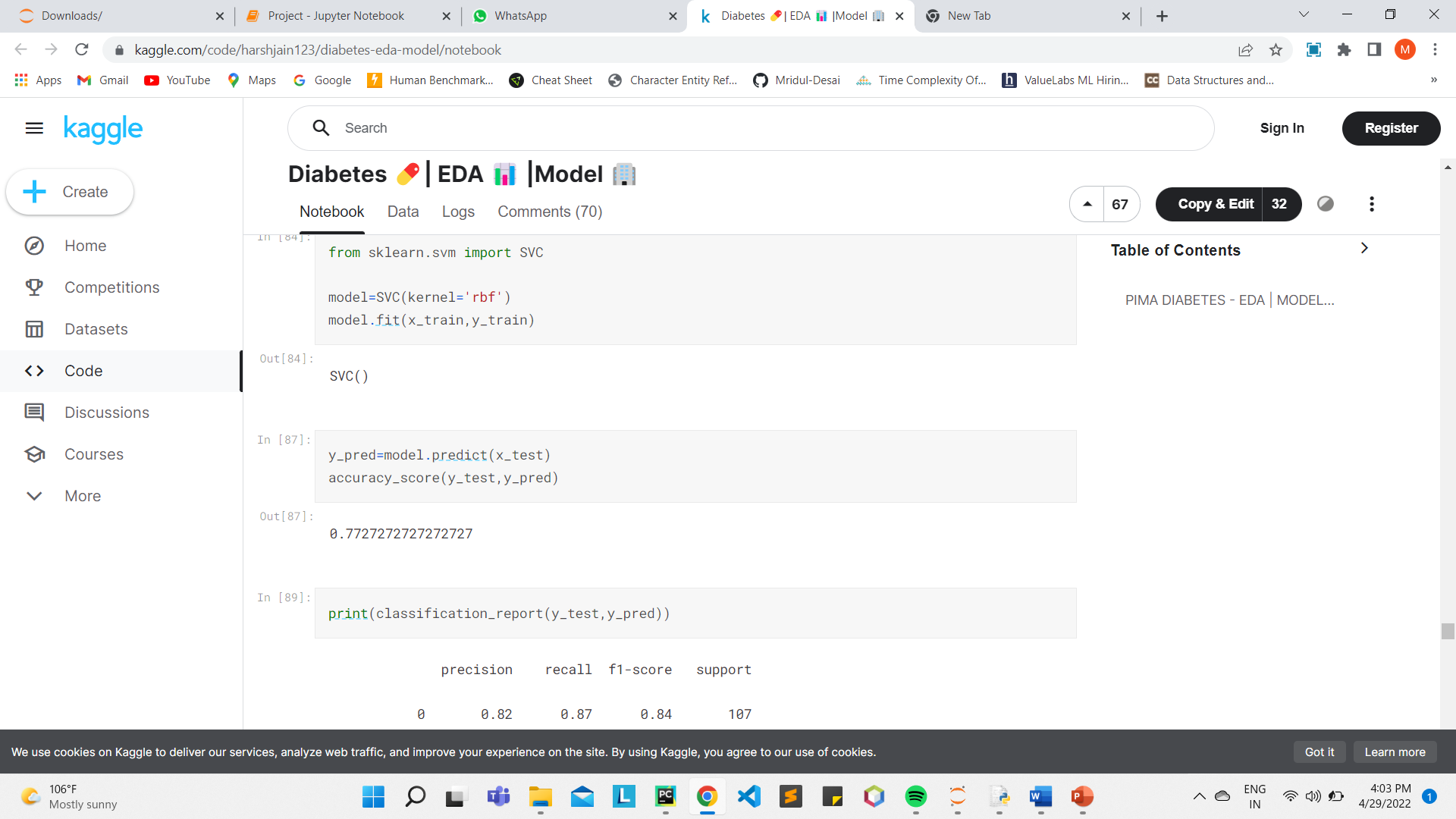


## 5.2 Test and Validation:

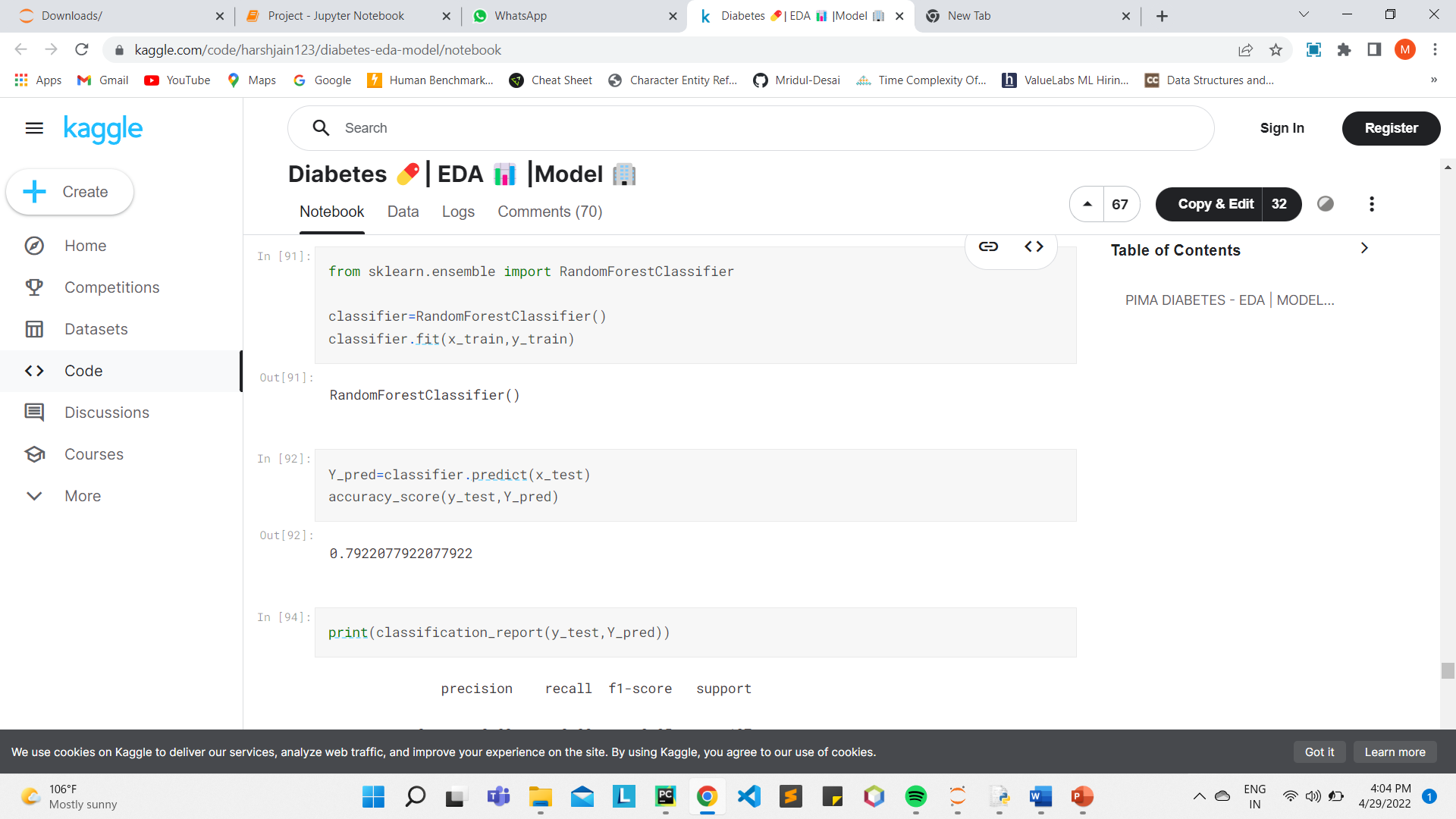
#### Feature Scaling

* Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.
* **Standard Scaler**: It is a very effective technique which re-scales a feature value so that it has distribution with 0 mean value and variance equals to 1.
* 

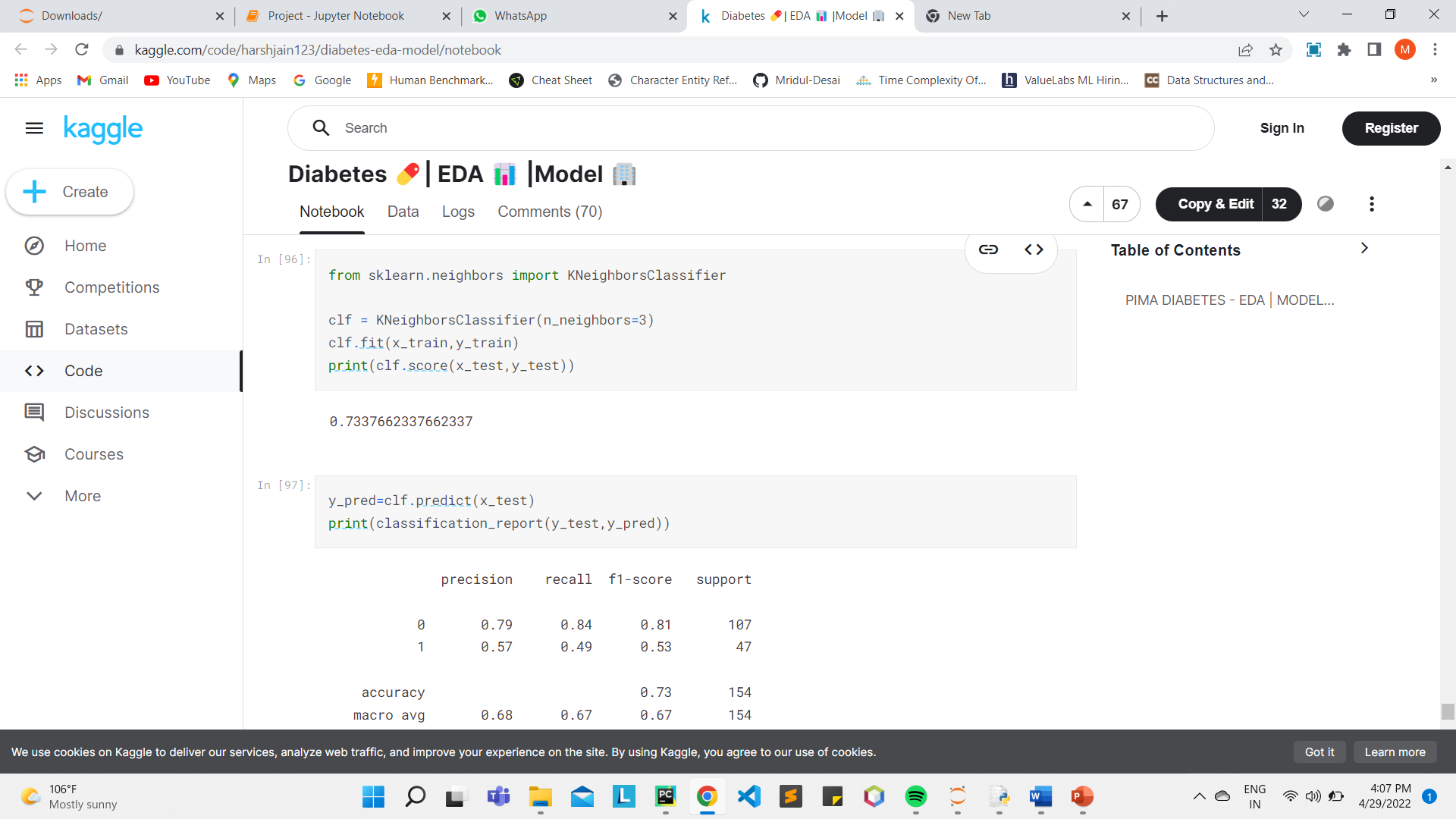
#### SVM Model

* Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.
* The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.
* 

#### Random Forest Model

* Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object.
* This works well because a single decision tree may be prone to a noise, but aggregate of many decision trees reduce the effect of noise giving more accurate results.
* 

#### KNN Model

* K-Nearest Neighbors is one of the simplest Machine Learning algorithms based on Supervised Learning technique. It can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
* K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
* 

### Conclusion

In this kernel, We have performed Exploratory Data Analysis, Data Preprocessing, Visualization of Features, Correlation Matrix, Model Building (SVM, RF, KNN).

**CHAPTER-6:**

**PROJECT OUTCOME AND APPLICABILITY**

# 6.1 Test and Validation

The recent trend in disease prediction technologies is the use of machine learning which makes predictions based on training of their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of KNN based Machine learning to predict disease. Factors considered are normal, low, high based on their occurrence in the respective ranges.

**6.2** **Key implementations outline of the System**

In this project “Stock Price Prediction” we discussed the design and implementation of Machine learning algorithm.The project is built using open-source software modules. The modular nature of this project makes it more flexible and easier to add additional features without disturbing current system functionalities.

**6.3 Significant project outcomes**

When new investors jump into the stock market, they need the support of something, like a professional trader, so that they cannot go with a loss. So here our project comes, which will provide the predicted data or the price of the stock from the old data, by using machine learning models. So, by using our project at least the investor would not make any loss and make some good profit.

**6.4 Project applicability on Real-world applications**

The development of e-Health mobile application using cloud environment has proved to be very useful in monitoring and managing patient’s clinical data. In recent years, there has been a proliferation of mobile health (mHealth) applications on smartphones.

mHealth applications are programs that use smartphone's inbuilt tools, such as the Global Positioning System (GPS), accelerometer, microphone, speaker, and camera to automatically detect and measure health-related behaviors. In order to measure and upload physiological data, a number of applications may also synchronize wirelessly with other wearable devices, such as wristband sensor, heart rate sensor, belt sensor, shoe sensor, glucometers, blood pressure cuffs, or smart clothing with wearable sensing technologies.

Most of the popular mobile health and fitness applications focus on self-monitoring physiological markers relevant to a person's health status and for encouraging physical activity and healthy diets.

Depending on the data collected, mHealth applications may cover a wide range of uses, including patient health and fitness self-management tools, remote and continuous monitoring of patients by caregivers, reminder systems improving glycemic control in patients with diabetes by prompting the users to take medications and check their blood glucose, symptom monitoring in asthma and heart disease, supporting smoking cessation, an effective source of health information, helping to monitor, improve, and manage fitness activities, weight goals, diet, pregnancy, and sleep.

They have the potential to reduce the cost of healthcare in a period in which many countries are facing the problem of an aging population with chronic diseases such as obesity, diabetes, and high pressure by encouraging healthy behaviors to prevent or reduce health problems, and by supporting chronic disease self-management that may reduce the number of healthcare visits.

**CHAPTER-7**

**CONCLUSIONS AND RECOMMENDATION**

## 7.1 Outline

## With the escalating spread of chronic diseases, the only way to overcome it is predicting it in advance, hence Disease Prediction plays a vital role in the future of healthcare.

## In present, there are several models to predict the diseases but they are less accurate. Here, we have proposed a model that uses KNN algorithm for the prediction of diseases that is more accurate and gives an accuracy of 98 percent. In the proposed method, we have used test data for our cause and used it as a dataset.

Hence, we have successfully implemented this cause of Disease Prediction.

## 7.2 Future Enhancements

We are going to modify our model to a more accurate and widely varied web application with more disease prediction options that will be of help to the users from every environment and nationality.

**Reference**

* 1. YouTube Tutorials
  2. Research Papers
* <https://www.researchgate.net/publication/347381005_Disease_Prediction_Using_Machine_Learning>
* <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3661426>