



RV Educational Institutions®  
RV College of Engineering®

Autonomous  
Institution Affiliated  
to Visvesvaraya  
Technological  
University, Belagavi

Approved by AICTE,  
New Delhi

*Go, change the world*

## IV Semester Assignment

### Report on

### Health Care

*Submitted in Partial fulfilment of the requirement*

*for the IV Semester MCA Academic requirements for the  
course*

**18MCA442**

### MASTER OF COMPUTER APPLICATIONS

**By**

USN	STUDENT NAME
1RV19MCA21	Chaitra B V
1RD19MCA24	Thejashree.R

Department of Master of Computer  
ApplicationsRV College of Engineering®,  
Mysuru Road

RV Vidyanikethan Post, Bengaluru –

560059June -2021

## Domain: Healthcare

### Topic: Heart Disease Prediction

#### **Problem Statement**

Heart Disease Prediction uses soft computing techniques to categorize whether a person has heart disease or not. The main problem of this project is to identify different parameters building soft computing model neural network in this case and training the model in such a way that accuracy is maximum. Building a final model and giving the relevant inputs as parameters and getting the output whether a person is likely to have heart disease or not.

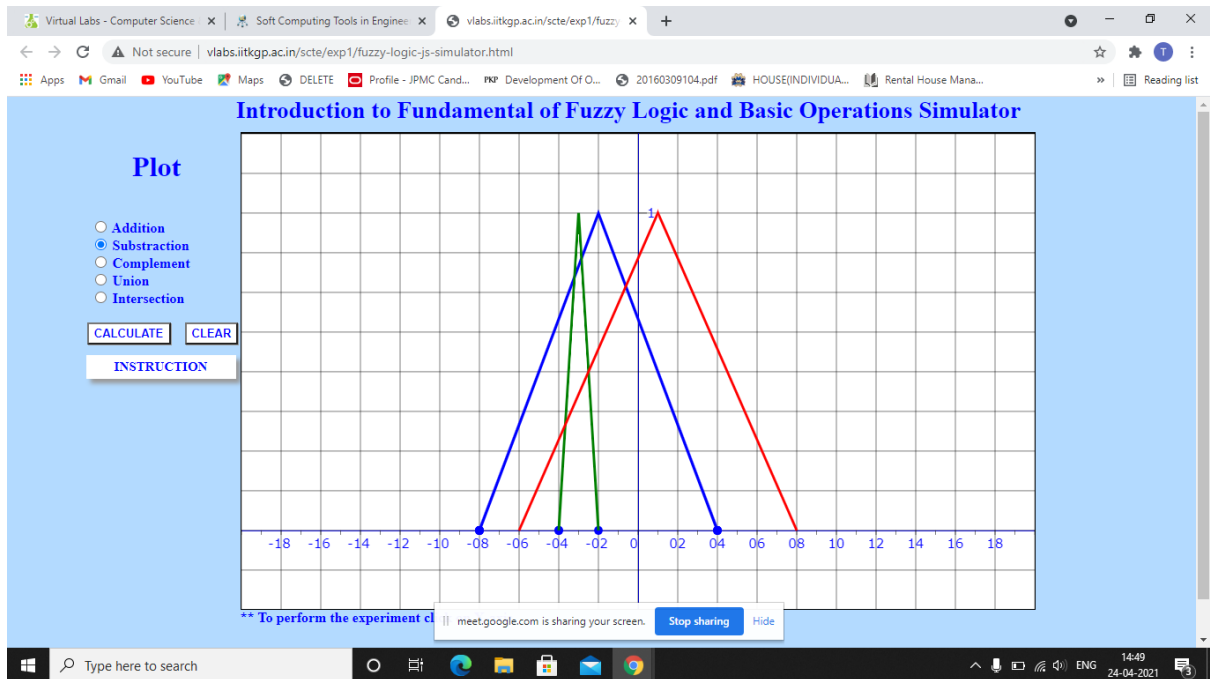
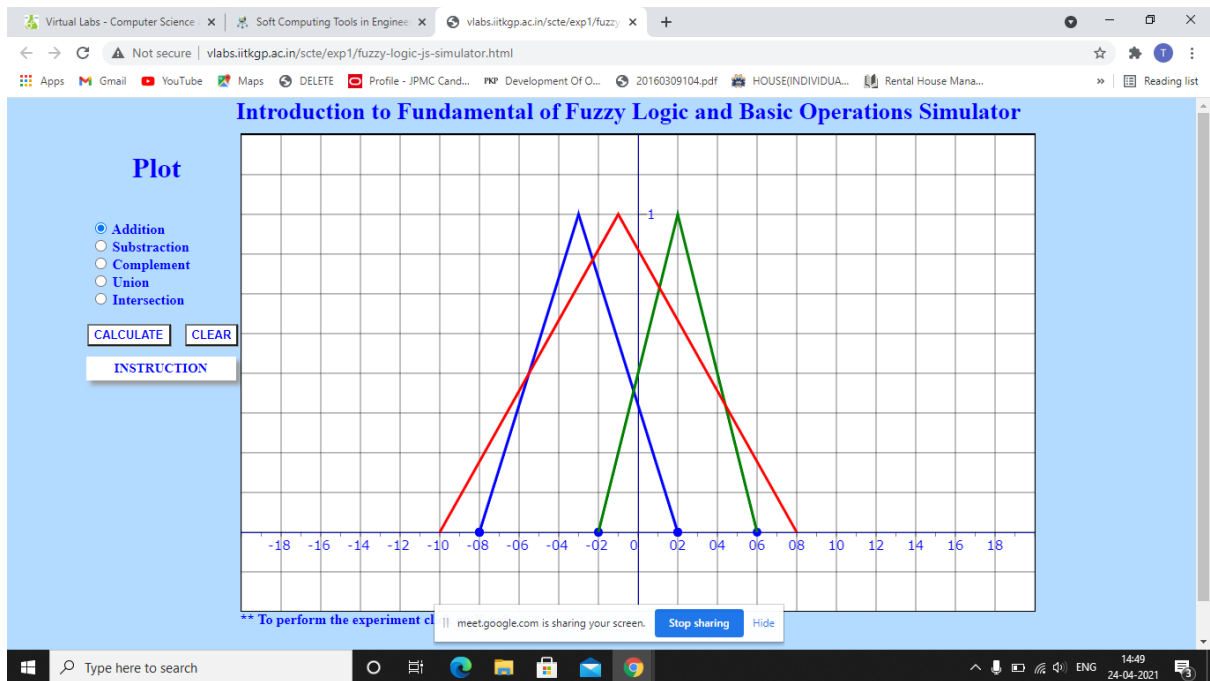
#### **Explanations and screen shots of the virtual lab program execution**

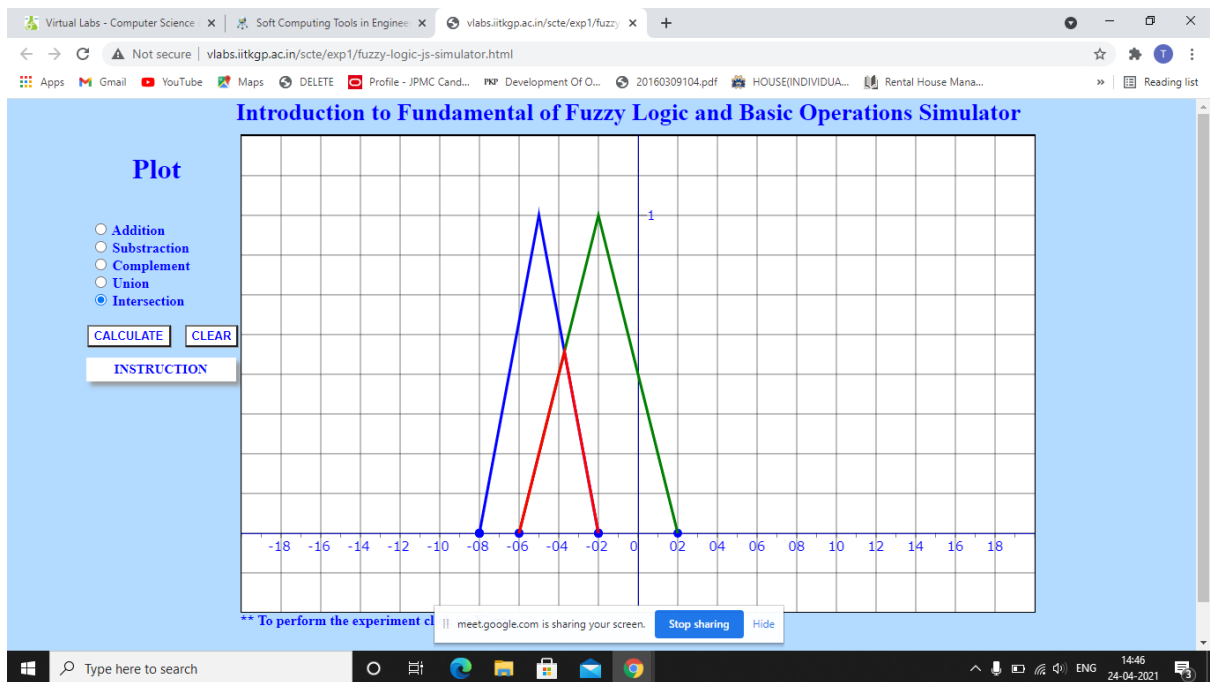
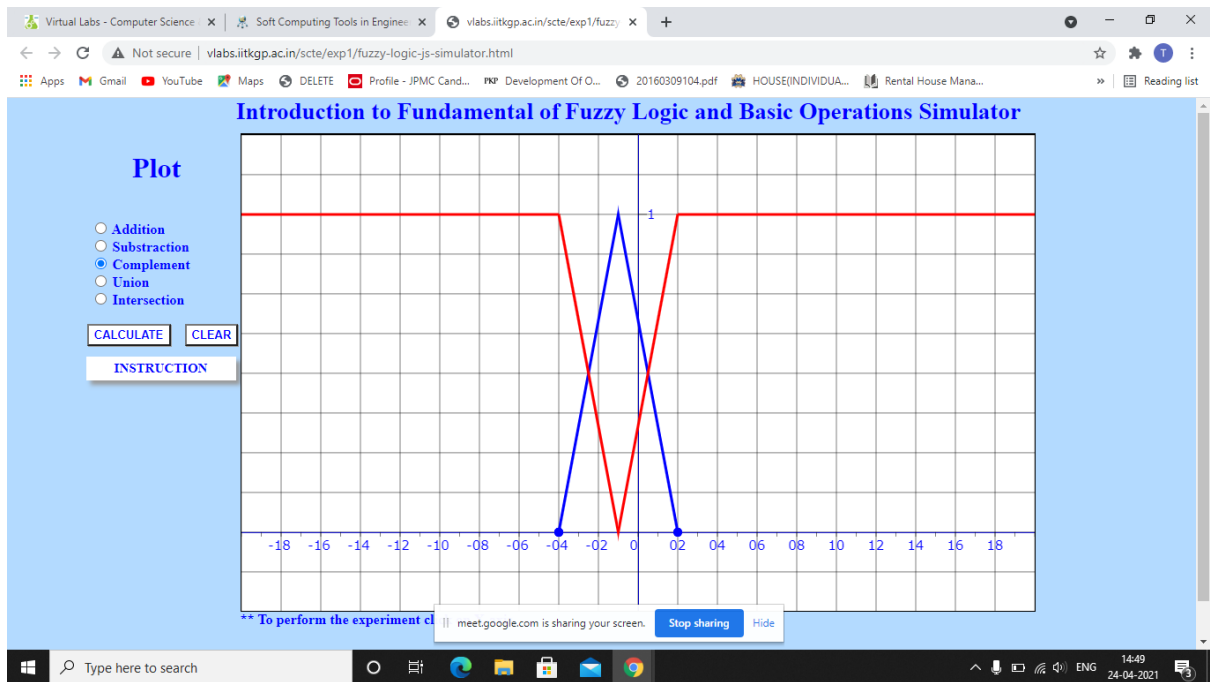
Virtual labs are simulated learning environments that allow students to complete laboratory experiments online and explore concepts and theories without stepping into a physical science lab. ... Virtual lab software creates opportunities for alternative access to science education.

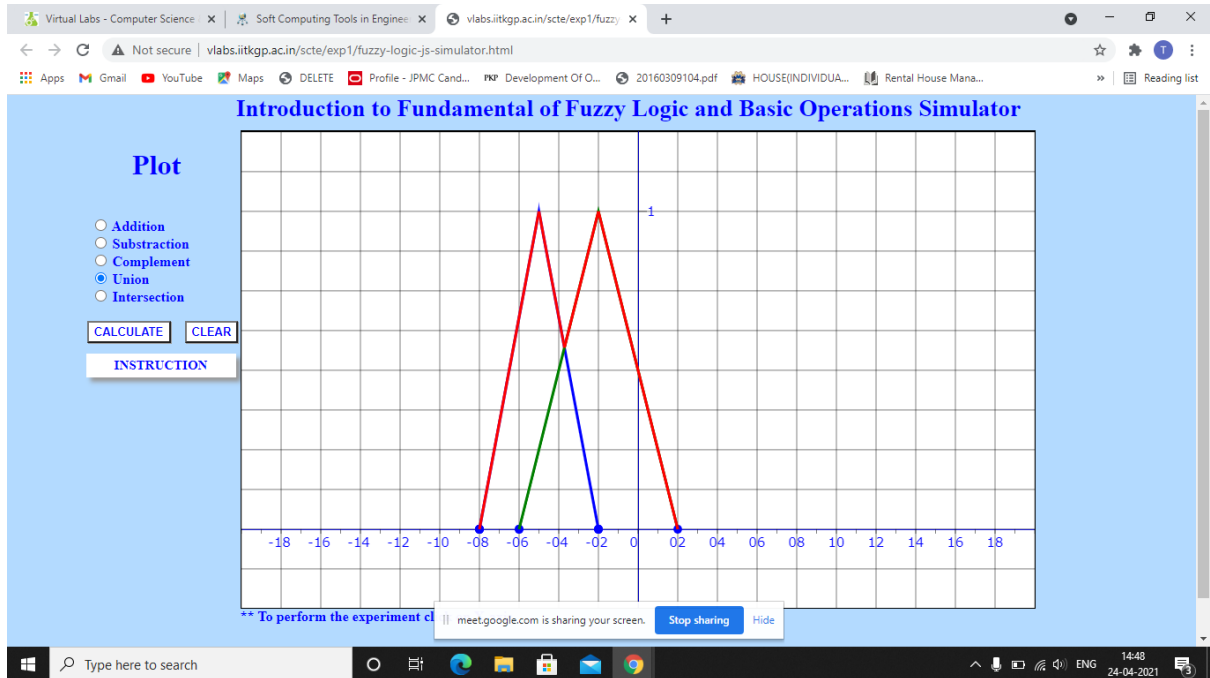
Fuzzy logic was developed by Lotfi A. Zadeh in the 1960s in order to provide mathematical rules and functions which permitted natural language queries. Fuzzy logic provides a means of calculating intermediate values between absolute true and absolute false with resulting values ranging between 0.0 and 1.0. With fuzzy logic

#### **Fuzzy Set Operations:**

- Fuzzy Addition
- Fuzzy Subtraction
- Fuzzy Complement
- Fuzzy Intersection
- Fuzzy Union



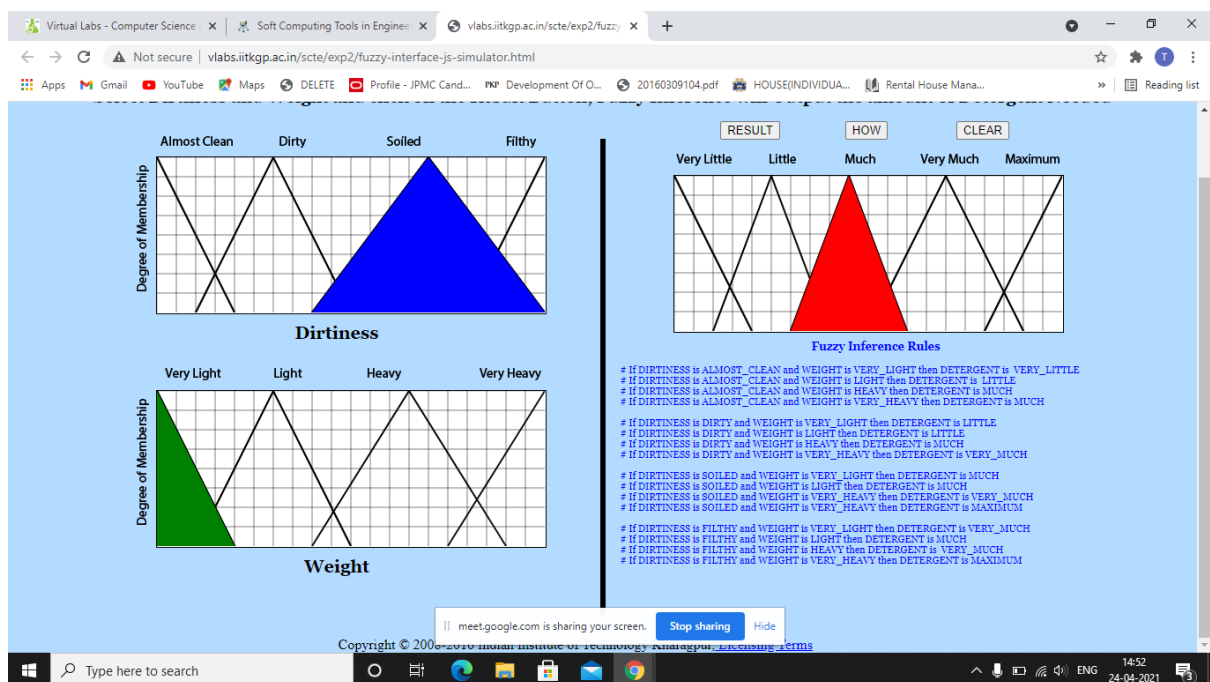


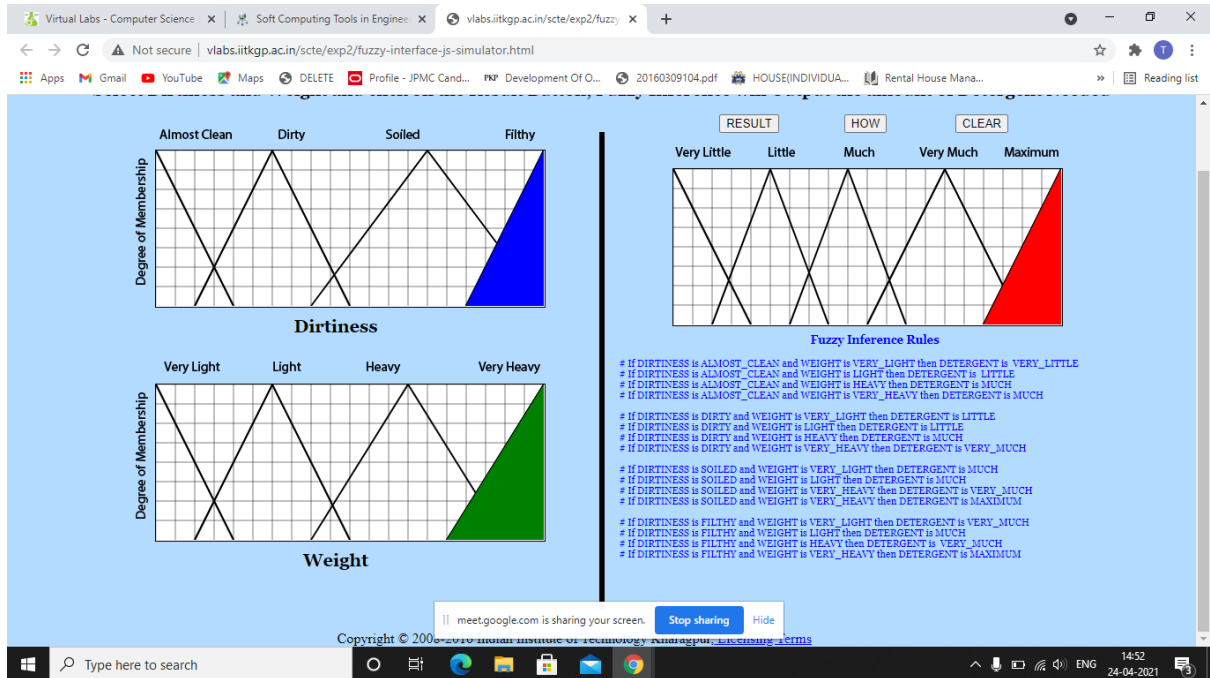
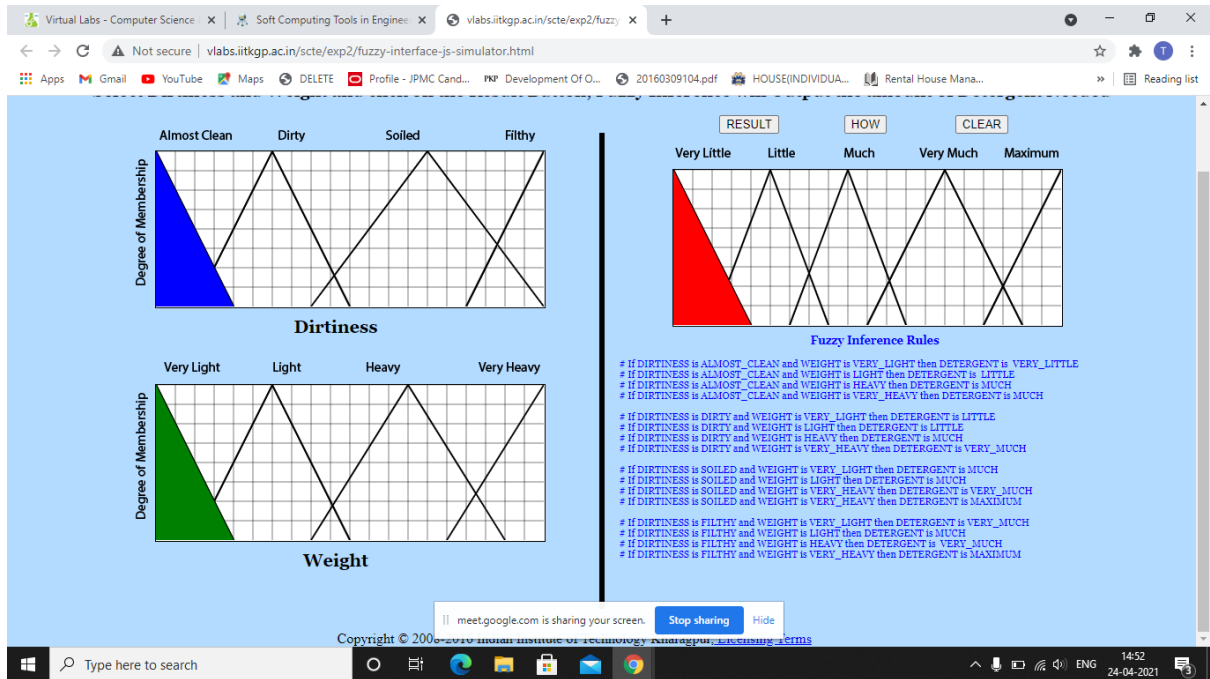


**Fuzzy inference** is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned.

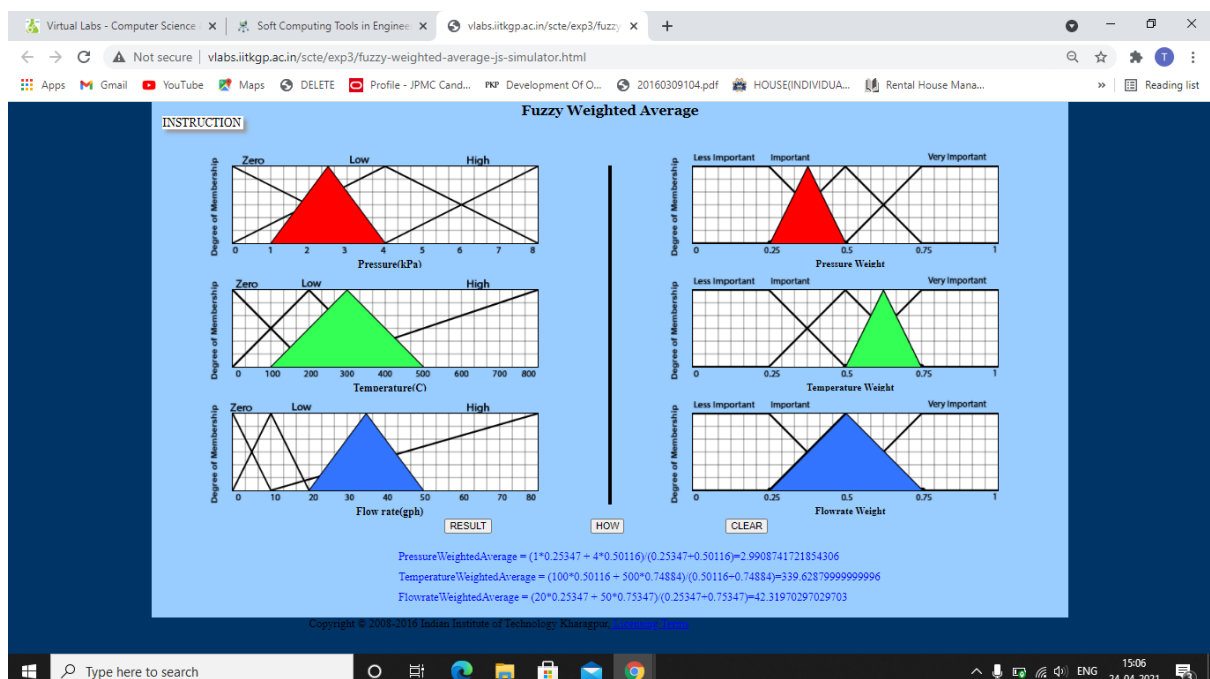
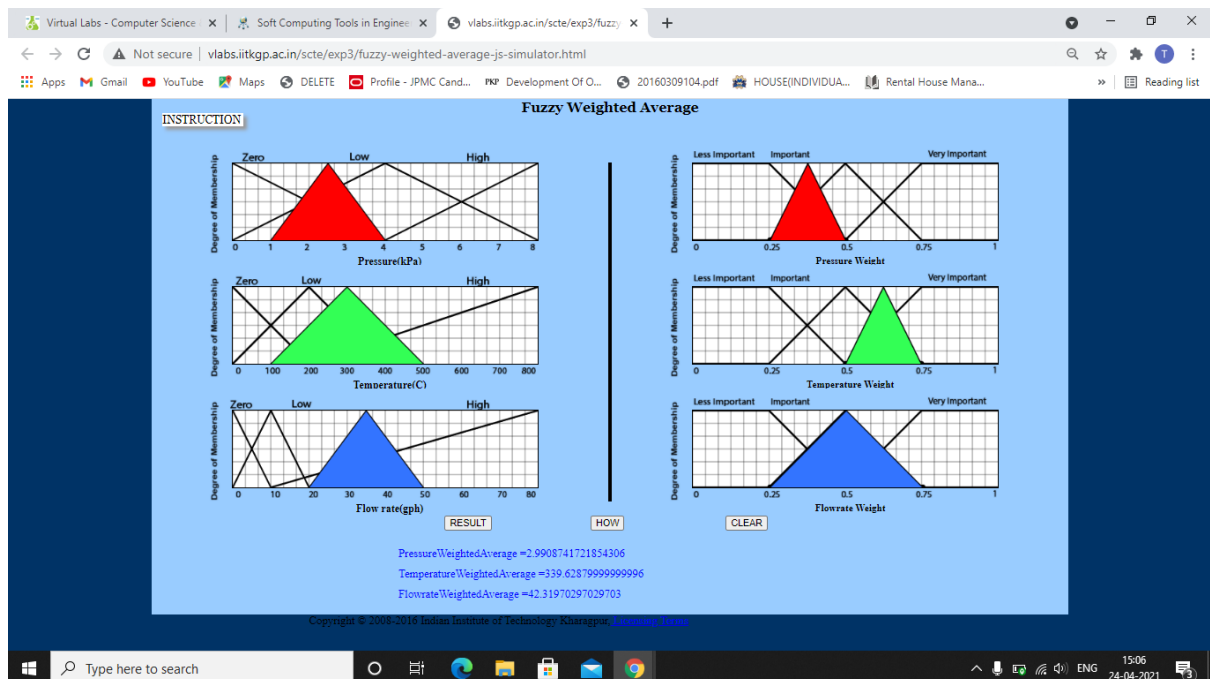
The process of fuzzy inference involves:

- Membership Functions
- Logical Operations
- If-Then Rules.





The fuzzy weighted average (FWA), which is a function of fuzzy numbers and is useful as an aggregation method in engineering or management science based on fuzzy sets theory. It provides a discrete approximate solution by  $\alpha$ -cuts level representation of fuzzy sets and interval analysis.





## Questions

1. Explain about the different fundamentals of soft computing approaches and platforms used in the identified domain? Justify the same?

Soft Computing is an emerging approach to computer like a human mind and to reason and learn in an environment of uncertainty and imprecision. It is the fusion of methodologies designed to model and enable solutions to real world problems.

The Fundamentals of soft computing are:

- Neural Networks
- Fuzzy Logic
- Genetic Algorithms

Neural Network:- Neural network is a interconnected network of large number of processing elements called neurons in an architecture inspired by the brain. It mimic the activities of the brain.

The main characteristics are:

- Mapping Capabilities/Pattern Association.
- Generalization
- Robustness
- Fault Tolerance
- Parallel and high speed information processing

Fuzzy Logic:- Fuzzy Logic set was proposed in 1965 by A.Zadeh. In classical set theory an element either belong to or does not belong to a set and hence such set are termed as crisp set. But in fuzzy set, many degrees of membership (between 0/1) are allowed.

Genetic Algorithm:- Genetic Algorithm is an algorithm which mimic some of the process of natural evolution. These are the stochastic search methods based on the principles of natural genetic system. They perform a search in providing an optimal solution for evaluation function of an optimization problem.

### Approach Used

The approach which is used in heart disease prediction is neural network. The neural network model is built to solve the problem of predicting whether a person is likely to have heart disease or not.

This model is built with the following process.

- Data Collection:-This project model is trained with the data set which I form which foundation.
- Feature Extraction:-The data set has several features extracting relevant features required for project and naming them.
- Training Neural Network Model:- using data sets with extracted features neural network model is trained.
- Result:- The neural network model predicts whether a person is likely to have heart diseases or not with relevant input.

## **2.Discuss the methodology adopted to solve the problem In the domain selected.**

The methodology adopted to solve the problem in heart disease prediction are:

- **Data Collection:-**Data set is the key element in building any models in soft computing. So the data set used in this project is Cleveland clinic Foundation heart disease set.
- **Data Preprocessing:-** After identifying required dataset the next step is data processing .It involves:
  - i. Missing values identification and removing them.
  - ii. Feature Extraction
  - iii. Splitting training and testing data.
- **Building Neural Network Model:-**This project uses koras package to build Neural Network Model.It has several dense layers input and output parameters and one hidden layer.
- **Training and Testing Data:-**After building Neural Network Model the next step is to fit the training data in it.
- **Building Classification:-**This step is for improving result. Binary classification is used to improve the result. Converting our categorical model which has result of [1,2,3,4] to binary model which has result of [0,1] heart disease or no hear disease.

### **3. Discuss and analyze the results obtained for the problem selected?**

The output/Result of the heart disease prediction is to determine whether the person is likely to have heart disease or not.

The project has two model i.e. categorical model and binary model. Binary model is built so that the accuracy of the model is maximum. Although categorical model results are promising there are large errors. This could be because it is difficult to distinguish the different to distinguish the different security levels of heart disease(1-4). So binary classification model is built which classifies(0-1) heart disease or no heart disease. By data exploration we can know that chest pain type, exercise induced angine, gt depression, slope, thalassemia are directly co-related with heart disease. This results can prove that heart diseases infect the old person and young person but probability of the old person is higher than the young person.

And we see that most people with heart disease have asymptomatic chest pain and usually the people who do not have heart disease have normal electrocardiographic, whereas the people who have heart disease have preferable left ventricular hypertrophy. People who have higher heart rate greater than 150 are more likely to have heart disease.

#### 4. Evaluate the solution obtained

The model is evaluated with accuracy results. Performance metrics is calculated using sklearn model

Accuracy score of the project is 0.83 for binary Neural Network model

---

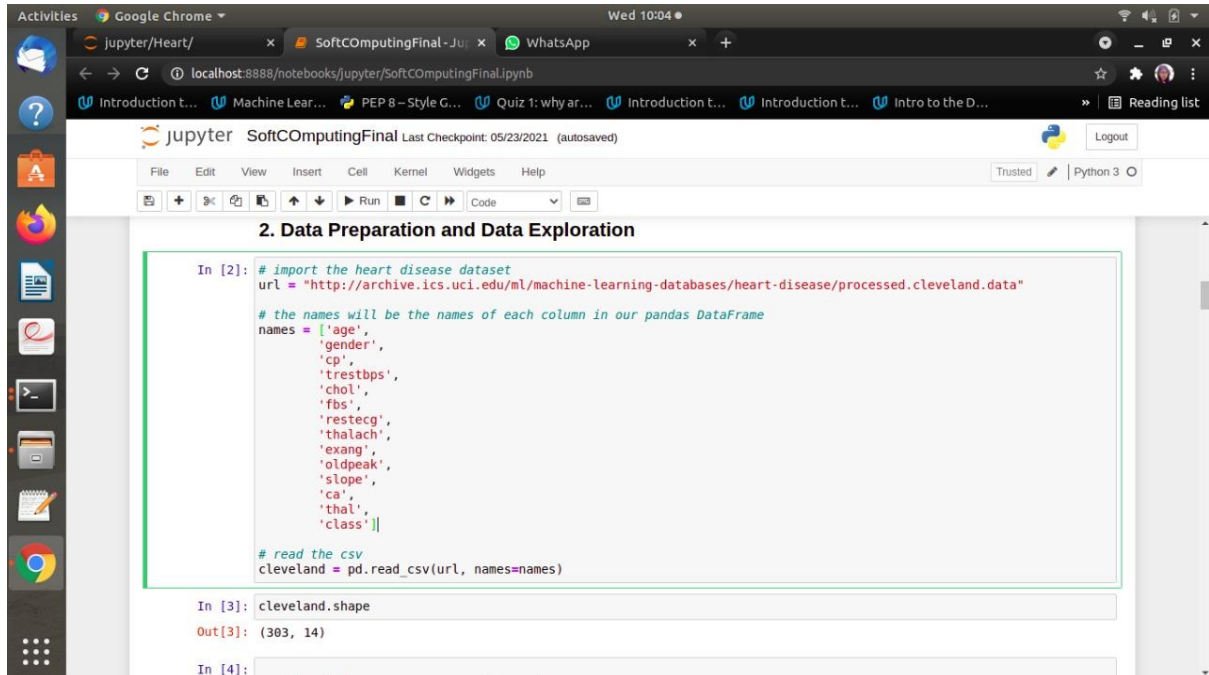
##### Results for Binary Model

0.8333333333333334

	precision	recall	f1-score	support
0	0.83	0.92	0.88	38
1	0.83	0.68	0.75	22
accuracy			0.83	60
macro avg	0.83	0.80	0.81	60
weighted avg	0.83	0.83	0.83	60

---

## Coding and Screenshots



The screenshot shows a Jupyter Notebook titled 'SoftComputingFinal' with a last checkpoint on 05/23/2021. The notebook is running on Python 3. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running, and code execution. The notebook content is as follows:

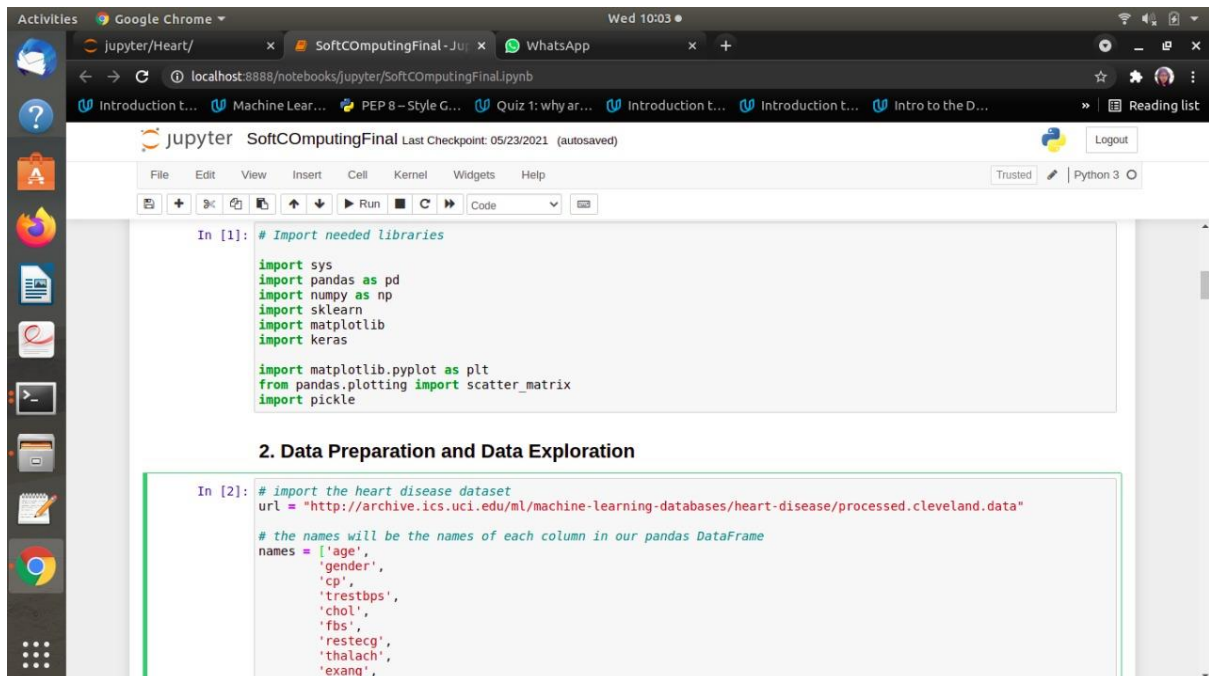
```
In [2]: # import the heart disease dataset
url = "http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data"

# the names will be the names of each column in our pandas DataFrame
names = ['age',
         'gender',
         'cp',
         'trestbps',
         'chol',
         'fbs',
         'restecg',
         'thalach',
         'exang',
         'oldpeak',
         'slope',
         'ca',
         'thal',
         'class']

# read the csv
cleveland = pd.read_csv(url, names=names)

In [3]: cleveland.shape
Out[3]: (303, 14)

In [4]:
```



The screenshot shows the same Jupyter Notebook interface, but with different code in the first cell. The notebook title and metadata remain the same. The code in the first cell is:

```
In [1]: # Import needed libraries

import sys
import pandas as pd
import numpy as np
import sklearn
import matplotlib
import keras

import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
import pickle
```

The second cell, titled '2. Data Preparation and Data Exploration', contains the same code as in the first screenshot:

```
In [2]: # import the heart disease dataset
url = "http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data"

# the names will be the names of each column in our pandas DataFrame
names = ['age',
         'gender',
         'cp',
         'trestbps',
         'chol',
         'fbs',
         'restecg',
         'thalach',
         'exang',
         'oldpeak',
         'slope',
         'ca',
         'thal',
         'class']

# read the csv
cleveland = pd.read_csv(url, names=names)
```

Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpynb

jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [5]: # remove missing data (indicated with a "?")
data = cleveland[~cleveland.isin(['?'])]
data.loc[280:]
```

281	47.0	1.0	3.0	130.0	253.0	0.0	0.0	179.0	0.0	0.0	1.0	0.0	3.0	0
282	55.0	0.0	4.0	128.0	205.0	0.0	1.0	130.0	1.0	2.0	2.0	1.0	7.0	3
283	35.0	1.0	2.0	122.0	192.0	0.0	0.0	174.0	0.0	0.0	1.0	0.0	3.0	0
284	61.0	1.0	4.0	148.0	203.0	0.0	0.0	161.0	0.0	0.0	1.0	1.0	7.0	2
285	58.0	1.0	4.0	114.0	318.0	0.0	1.0	140.0	0.0	4.4	3.0	3.0	6.0	4
286	58.0	0.0	4.0	170.0	225.0	1.0	2.0	146.0	1.0	2.8	2.0	2.0	6.0	2
287	58.0	1.0	2.0	125.0	220.0	0.0	0.0	144.0	0.0	0.4	2.0	NaN	7.0	0
288	56.0	1.0	2.0	130.0	221.0	0.0	2.0	163.0	0.0	0.0	1.0	0.0	7.0	0
289	56.0	1.0	2.0	120.0	240.0	0.0	0.0	169.0	0.0	0.0	3.0	0.0	3.0	0
290	67.0	1.0	3.0	152.0	212.0	0.0	2.0	150.0	0.0	0.8	2.0	0.0	7.0	1
291	55.0	0.0	2.0	132.0	342.0	0.0	0.0	166.0	0.0	1.2	1.0	0.0	3.0	0
292	44.0	1.0	4.0	120.0	169.0	0.0	0.0	144.0	1.0	2.8	3.0	0.0	6.0	2
293	63.0	1.0	4.0	140.0	187.0	0.0	2.0	144.0	1.0	4.0	1.0	2.0	7.0	2

```
In [6]: # drop rows with NaN values from DataFrame
data = data.dropna(axis=0)
data.loc[280:]
```

Out[6]:

Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpynb

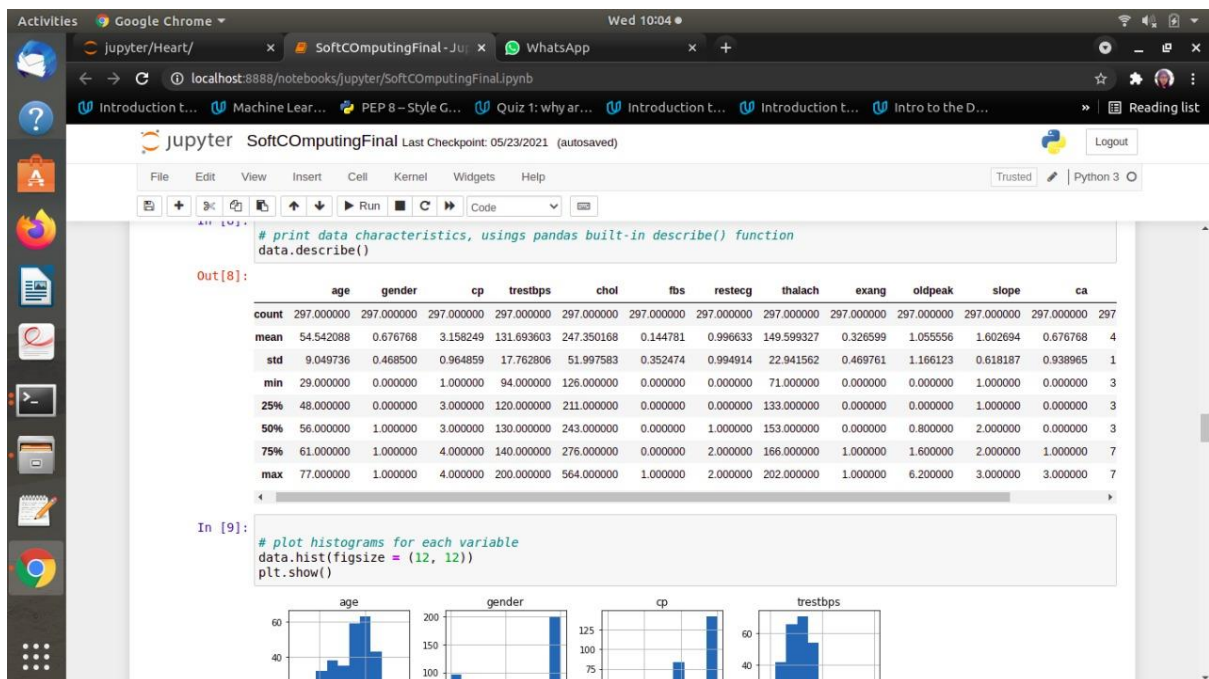
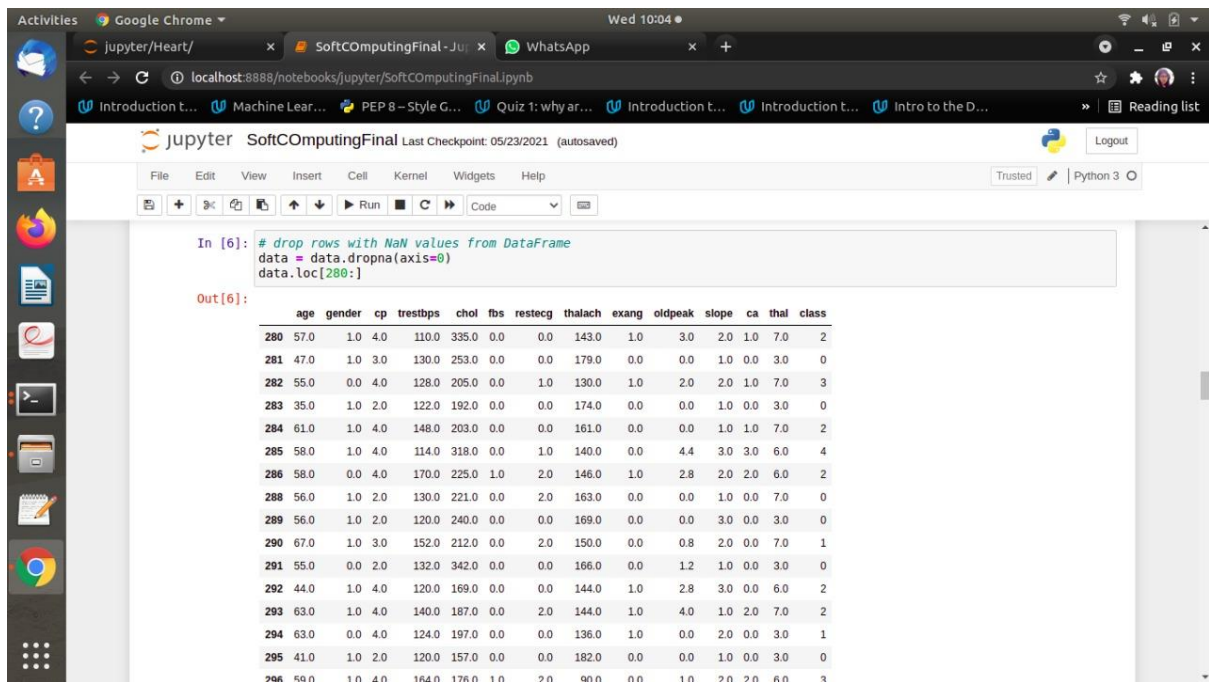
jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

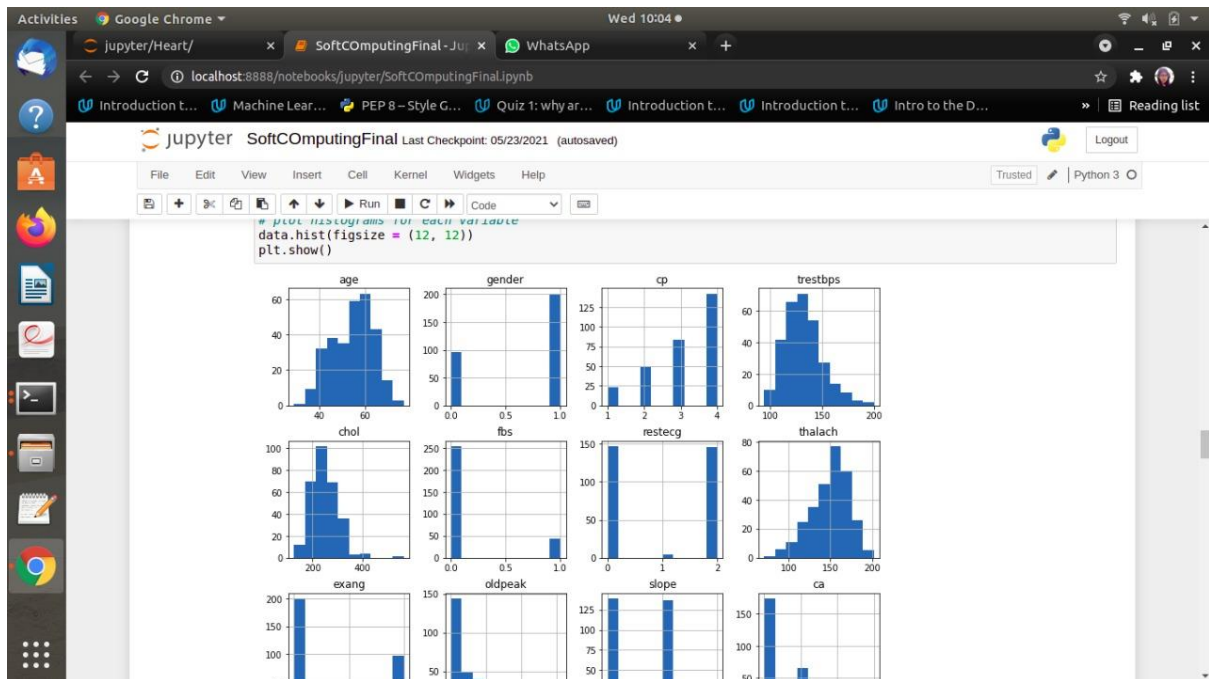
```
In [4]: # print the last twenty or so data points
cleveland.loc[280:]
```

Out[4]:

	age	gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	class
280	57.0	1.0	4.0	110.0	335.0	0.0	0.0	143.0	1.0	3.0	2.0	1.0	7.0	2
281	47.0	1.0	3.0	130.0	253.0	0.0	0.0	179.0	0.0	0.0	1.0	0.0	3.0	0
282	55.0	0.0	4.0	128.0	205.0	0.0	1.0	130.0	1.0	2.0	2.0	1.0	7.0	3
283	35.0	1.0	2.0	122.0	192.0	0.0	0.0	174.0	0.0	0.0	1.0	0.0	3.0	0
284	61.0	1.0	4.0	148.0	203.0	0.0	0.0	161.0	0.0	0.0	1.0	1.0	7.0	2
285	58.0	1.0	4.0	114.0	318.0	0.0	1.0	140.0	0.0	4.4	3.0	3.0	6.0	4
286	58.0	0.0	4.0	170.0	225.0	1.0	2.0	146.0	1.0	2.8	2.0	2.0	6.0	2
287	58.0	1.0	2.0	125.0	220.0	0.0	0.0	144.0	0.0	0.4	2.0	?	7.0	0
288	56.0	1.0	2.0	130.0	221.0	0.0	2.0	163.0	0.0	0.0	1.0	0.0	7.0	0
289	56.0	1.0	2.0	120.0	240.0	0.0	0.0	169.0	0.0	0.0	3.0	0.0	3.0	0
290	67.0	1.0	3.0	152.0	212.0	0.0	2.0	150.0	0.0	0.8	2.0	0.0	7.0	1
291	55.0	0.0	2.0	132.0	342.0	0.0	0.0	166.0	0.0	1.2	1.0	0.0	3.0	0
292	44.0	1.0	4.0	120.0	169.0	0.0	0.0	144.0	1.0	2.8	3.0	0.0	6.0	2
293	63.0	1.0	4.0	140.0	187.0	0.0	2.0	144.0	1.0	4.0	1.0	2.0	7.0	2
294	63.0	0.0	4.0	124.0	197.0	0.0	0.0	136.0	1.0	0.0	2.0	0.0	3.0	1
295	41.0	1.0	2.0	120.0	157.0	0.0	0.0	187.0	0.0	0.0	1.0	0.0	2.0	0







Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpynb

Introduction t... Machine Lear... PEP 8 - Style G... Quiz 1: why ar... Introduction t... Introduction t... Intro to the D... Reading list

Jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [10]: # create X and Y datasets for training
from sklearn import model_selection

X = np.array(data.drop(['class'], 1))
y = np.array(data['class'])

X_train, X_test, y_train, y_test = model_selection.train_test_split(X, y, test_size = 0.2)

In [11]: # convert the data to categorical labels
from keras.utils.np_utils import to_categorical

Y_train = to_categorical(y_train, num_classes=None)
Y_test = to_categorical(y_test, num_classes=None)
print(Y_train)

[[0. 0. 0. 0. 1.]
 [0. 0. 0. 1. 0.]
 [1. 0. 0. 0. 0.]
 ...
 [0. 0. 0. 1. 0.]
 [0. 0. 0. 1. 0.]
 [0. 0. 1. 0. 0.]
 [0. 0. 1. 0. 0.]]
```

**3. Building Neural Network Model**

Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpynb

jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

### 3. Building Neural Network Model

```
In [12]: import tensorflow as tf
from keras.models import Sequential
from keras.layers import Dense

# define a function to build the keras model
def create_model():
    # create model
    model = Sequential()
    model.add(Dense(8, input_dim=13, kernel_initializer='normal', activation='relu'))
    model.add(Dense(4, kernel_initializer='normal', activation='relu'))
    model.add(Dense(5, activation='softmax'))

    model.compile(loss='categorical_crossentropy', metrics=['accuracy'])
    return model

model = create_model()
```

### 4. Training and testing Data

```
In [13]: # fit the model to the training data
```

Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpynb

jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

### 4. Training and testing Data

```
In [13]: # fit the model to the training data
model.fit(X_train, Y_train, epochs=100, batch_size=10, verbose = 1)
```

```
Epoch 0/100
24/24 [=====] - 0s 2ms/step - loss: 1.2113 - accuracy: 0.5243
Epoch 9/100
24/24 [=====] - 0s 2ms/step - loss: 1.2170 - accuracy: 0.5348
Epoch 10/100
24/24 [=====] - 0s 2ms/step - loss: 1.2879 - accuracy: 0.4904
Epoch 11/100
24/24 [=====] - 0s 3ms/step - loss: 1.3329 - accuracy: 0.4609
Epoch 12/100
24/24 [=====] - 0s 3ms/step - loss: 1.1371 - accuracy: 0.5524
Epoch 13/100
24/24 [=====] - 0s 3ms/step - loss: 1.2530 - accuracy: 0.4791
Epoch 14/100
24/24 [=====] - 0s 2ms/step - loss: 1.2559 - accuracy: 0.4797
Epoch 15/100
24/24 [=====] - 0s 3ms/step - loss: 1.2637 - accuracy: 0.4991
Epoch 16/100
24/24 [=====] - 0s 2ms/step - loss: 1.1111 - accuracy: 0.5632
Epoch 17/100
24/24 [=====] - 0s 2ms/step - loss: 1.2243 - accuracy: 0.5045
```

```
In [14]: # convert into binary classification problem - heart disease or no heart disease
Y_train_binary = Y_train.copy()
Y_test_binary = Y_test.copy()
```

Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpyb

jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [14]: # convert into binary classification problem - heart disease or no heart disease
Y_train_binary = y_train.copy()
Y_test_binary = y_test.copy()

Y_train_binary[Y_train_binary > 0] = 1
Y_test_binary[Y_test_binary > 0] = 1

In [23]: # define a new keras model for binary classification
def create_binary_model():
    # create model
    model = Sequential()
    model.add(Dense(8, input_dim=13, kernel_initializer='normal', activation='relu'))
    model.add(Dense(4, kernel_initializer='normal', activation='relu'))
    model.add(Dense(1, activation='sigmoid'))

    # Compile model
    #adam = Adam(lr=0.001)
    model.compile(loss='binary_crossentropy', metrics=['accuracy'])
    return model

binary_model = create_binary_model()

In [24]: # fit the binary model on the training data
binary_model.fit(X_train, Y_train_binary, epochs=100, batch_size=10, verbose = 1)
```

Activities Google Chrome Wed 10:04

localhost:8888/notebooks/jupyter/SoftComputingFinalIpyb

jupyter SoftComputingFinal Last Checkpoint: 05/23/2021 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [25]: # generate classification report using predictions for binary model
binary_pred = np.round(binary_model.predict(X_test)).astype(int)

print('Results for Binary Model')
print(accuracy_score(Y_test_binary, binary_pred))
print(classification_report(Y_test_binary, binary_pred))

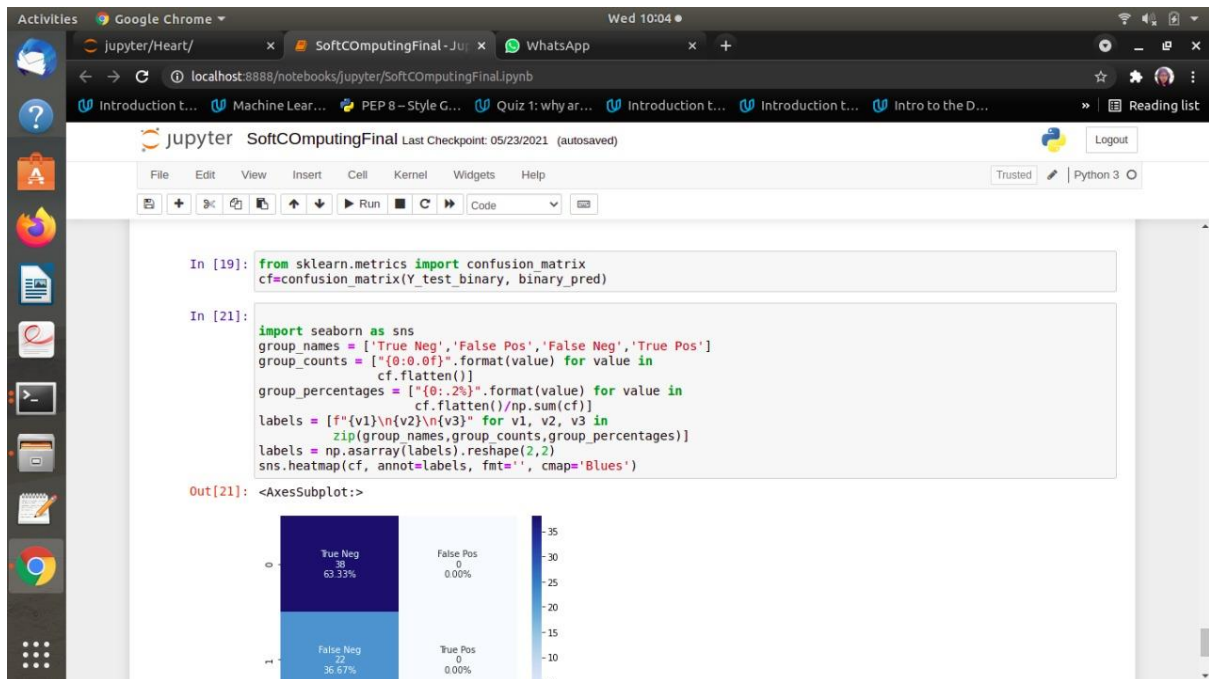
Results for Binary Model
0.8333333333333334
          precision    recall  f1-score   support

     0       0.83        0.92        0.88         38
     1       0.83        0.68        0.75         22

 accuracy          0.83
 macro avg         0.83        0.80        0.81         60
 weighted avg      0.83        0.83        0.83         60

In [19]: from sklearn.metrics import confusion_matrix
cf=confusion_matrix(Y_test_binary, binary_pred)

In [21]: import seaborn as sns
group_names = ['True Neg', 'False Pos', 'False Neg', 'True Pos']
group_counts = ["{0:0.0f}".format(value) for value in
                cf.flatten()]
group_percentages = ["{0:.2%}".format(value) for value in
```



Activities Google Chrome Wed 10:02

Jupyter/Heart/ x SoftComputingFinal - Ju x Heart Disease Test x +

127.0.0.1:5000

Introduction t... Machine Lear... PEP 8 - Style G... Quiz 1: why ar... Introduction t... Introduction t... Intro to the D... Reading list

### Heart Disease Test Form

Age	Sex		
58	Male		

Chest Pain Type	Resting Blood Pressure in mm Hg	Serum Cholesterol in mg/dl	Fasting Blood Sugar > 120 mg/dl
Typical Angina	90	250	False

Resting ECG Results	Maximum Heart Rate	ST Depression Induced	Exercise Induced Angina
Normal	80	4	No

Slope of the Peak Exercise ST Segment	Number of Vessels Colored by Fluoroscopy	Thalassemia
Upsloping	2	Normal

**Result**

Activities Google Chrome Wed 10:01

Jupyter/Heart/ x SoftComputingFinal - Ju x Heart Disease Test x +

127.0.0.1:5000/predict

Introduction t... Machine Lear... PEP 8 - Style G... Quiz 1: why ar... Introduction t... Introduction t... Intro to the D... Reading list

### Heart Disease Test Form

Age	Sex		
	-- Select an Option --		

Chest Pain Type	Resting Blood Pressure in mm Hg	Serum Cholesterol in mg/dl	Fasting Blood Sugar > 120 mg/dl
-- Select an Option --			-- Select an Option --

Resting ECG Results	Maximum Heart Rate	ST Depression Induced	Exercise Induced Angina
-- Select an Option --			-- Select an Option --

Slope of the Peak Exercise ST Segment	Number of Vessels Colored by Fluoroscopy	Thalassemia
-- Select an Option --	-- Select an Option --	-- Select an Option --

**Result**

**The patient is likely to have heart disease!**



Activities Google Chrome Wed 10:01

jupyter/Heart/ x SoftCOMputingFinal - Ju x Heart Disease Test x +

127.0.0.1:5000

Introduction t... Machine Lear... PEP 8 - Style G... Quiz 1: why ar... Introduction t... Introduction t... Intro to the D... Reading list

### Heart Disease Test Form

Age	Sex		
58	Male		
Chest Pain Type	Resting Blood Pressure in mm Hg	Serum Cholestoral in mg/dl	Fasting Blood Sugar > 120 mg/dl
Asymptomatic	114	318	False
Resting ECG Results	Maximum Heart Rate	ST Depression Induced	Exercise Induced Angina
Having ST-T wave abnormal	140	4	No
Slope of the Peak Exercise ST Segment	Number of Vessels Colored by Flourosopy	Thalassemia	
Downsloping	3	Fixed defect	

Result

Activities Google Chrome Wed 10:03

jupyter/Heart/ x SoftCOMputingFinal - Ju x Heart Disease Test x +

127.0.0.1:5000/predict

Introduction t... Machine Lear... PEP 8 - Style G... Quiz 1: why ar... Introduction t... Introduction t... Intro to the D... Reading list

### Heart Disease Test Form

Age	Sex		
	-- Select an Option --		
Chest Pain Type	Resting Blood Pressure in mm Hg	Serum Cholestoral in mg/dl	Fasting Blood Sugar > 120 mg/dl
-- Select an Option --			-- Select an Option --
Resting ECG Results	Maximum Heart Rate	ST Depression Induced	Exercise Induced Angina
-- Select an Option --			-- Select an Option --
Slope of the Peak Exercise ST Segment	Number of Vessels Colored by Flourosopy	Thalassemia	
-- Select an Option --	-- Select an Option --	-- Select an Option --	

Result

The patient is not likely to have heart disease!