A MINI PROJECT REPORT

ON

**“Disease Prediction Using Machine Learning”**

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF

BACHELOR OF COMPUTER ENGINEERING

BY

**Rohan Bordekar**

**Abhishek Budar**

**Shubham Gavali**

**Rohit Ghumare**

UNDER GUIDANCE OF

**Raju Mende**



**UNIVERSITY OF MUMBAI**

**DEPARTMENT OF COMPUTER ENGINEERING**

**DATTA MEGHE COLLEGE OF ENGINEERING**

*PLOT NO.98 SECTOR-3, AIROLI, NAVI MUMBAI*

**ACADAMIC YEAR 2018-19**

****

**DATTA MEGHE COLLEGE OF ENGINEERING**

**AIROLI, NAVI MUMBAI**

**CERTIFICATE**

This is to certify that the project entitled “**Disease Prediction Using Machine Learning**” is bona fide work of “**Rohan Bordekar,Abhishek Budar,Shubham Gavali,Rohit Ghumare**“ submitted to the University of Mumbai in partial fulfilment of the requirement for the award of **TE** in **“Computer Engineering”**.

Raju Mende Prof. A.P.Pande Dr.S.D.Sawarkar

Project Guide Head of the Department Principal

****

**DATTA MEGHE COLLEGE OF ENGINEERING**

**AIROLI, NAVI MUMBAI**

**MINI PROJECT APPROVAL**

This project report entitled “**Disease Prediction Using Machine Learning**” of the students“**Rohan Bordekar,Abhishek Budar,Shubham Gavali,Rohit Ghumare”**approved for the degree of **Computer Engineering.**

Internal Examiner External Examiner

Date: Date:

Place: Airoli Place: Airoli

**ACKNOWLEDGEMENT**

Motivation and guidance are the keys towards success. I would like to extend my thanks to all the sources of motivation.

We would like to grab this opportunity to thank **Dr.S.D.Sawarkar, Principal** for encouragement and support he has given for our project.

We express our deep gratitude to **Prof. A.P.Pande, Head of the Department** who has been the constant driving force behind the completion of this project.

We wish to express our heartfelt appreciation and deep sense of gratitude to my project guide **Prof. Raju Mende** for his encouragement, invaluable support, timely help, lucid suggestions and excellent guidance which helped us to understand and achieve the project goal. His concrete directions and critical views have greatly helped us in successful completion of this work.

We extend our sincere appreciation to all Professors for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individually.

We are also thankful to all those who helped us directly or indirectly in completion of this work.

Place: Airoli **Rohan Bordekar,Abhishek Budar,Shubham Gavali,Rohit Ghumare**

Date:

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr.No | CONTENTS | | Page No. |
| 1 | | INTRODUCTION | 1 |
| 2 | | REQUIREMENT ANALYSIS | 1 |
| 3 | | TECHNOLOGIES USED | 1 |
| 4 | | PROGRAM DESIGN (FLOWCHART) | 2 |
| 5 | | SOURCE CODE | 2 |
| 6 | | RESULTS & DISCUSSIONS | 9 |
| 7 | | CONCLUSION AND FUTURE SCOPE | 10 |

***INTRODUCTION:***

Healthcare industry has become big business. The healthcare industry produces large amounts of health-care data daily that can be used to extract information for predicting disease that can happen to a patient in future while using the treatment history and health data. This hidden information in the healthcare data will be later used for affective decision making for patient’s health. Also, this area need improvement by using the informative data in healthcare. Major challenge is how to extract the information from these data because the amount is very large so some data mining and machine learning techniques can be used. Also, the expected outcome and scope of this project is that if disease can be predicted than early treatment can be given to the patients which can reduce the risk of life and save life of patients and cost to get treatment of diseases can be reduced up to some extent by early recognition.

Machine Learning can play an essential role in predicting presence/absence of disorders, Heart diseases and more. Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment per patient basis.

In this article, We shall discuss a project where we worked on predicting potential Diseases in people using Machine Learning algorithms.

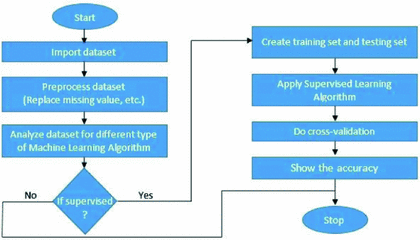
***REQUIREMENT ANALYSIS:***

The requirement of this project should be able to approximate a disease from the given data using linear regression.

***TECHNOLOGIES USED:***

1. Python.
2. CSV.
3. Numpy.
4. Pandas
5. Scikit-Learn
6. Tkinter

***FLOW CHART :***



***SOURCE CODE:***

from tkinter import \*

import numpy as np

import pandas as pd

# from gui\_stuff import \*

l1=['back\_pain','constipation','abdominal\_pain','diarrhoea','mild\_fever','yellow\_urine',

'yellowing\_of\_eyes','acute\_liver\_failure','fluid\_overload','swelling\_of\_stomach',

'swelled\_lymph\_nodes','malaise','blurred\_and\_distorted\_vision','phlegm','throat\_irritation',

'redness\_of\_eyes','sinus\_pressure','runny\_nose','congestion','chest\_pain','weakness\_in\_limbs',

'fast\_heart\_rate','pain\_during\_bowel\_movements','pain\_in\_anal\_region','bloody\_stool',

'irritation\_in\_anus','neck\_pain','dizziness','cramps','bruising','obesity','unsteadiness',

'weakness\_of\_one\_body\_side','loss\_of\_smell','bladder\_discomfort','foul\_smell\_of urine',

'continuous\_feel\_of\_urine','passage\_of\_gases','internal\_itching','toxic\_look\_(typhos)',

'depression','irritability','muscle\_pain','altered\_sensorium','red\_spots\_over\_body','belly\_pain',

'blister','red\_sore\_around\_nose','yellow\_crust\_ooze']

disease=['Fungal infection','Allergy','GERD','Chronic cholestasis','Drug Reaction',

'Peptic ulcer diseae','AIDS','Diabetes','Gastroenteritis','Bronchial Asthma','Hypertension',

' Migraine','Cervical spondylosis','Paralysis (brain hemorrhage)','Jaundice','Malaria','Chicken

pox','Dengue','Typhoid','hepatitisA','CommonCold','Pneumonia','Dimorphichemmorhoids(pile

s)','Heartattack','Varicoseveins','Hypothyroidism','Hyperthyroidism','Hypoglycemia','Osteoarthristis',]

l2=[]

for x in range(0,len(l1)):

l2.append(0)

df=pd.read\_csv("Training.csv")

df.replace({'prognosis':{'Fungalinfection':0,'Allergy':1,'GERD':2,'Chronic cholestasis':3,'Drug Reaction':4, 'Peptic ulcer diseae':5,'AIDS':6,'Diabetes ':7,'Gastroenteritis':8,'Bronchial Asthma':9,'Hypertension':10,'Migraine':11,'Cervicalspondylosis':12,'Paralysis(brain hemorrhage)':13,'Jaundice':14,'Malaria':15,'Chickenpox':16,'Dengue':17,'Typhoid':18,'hepatitisA':19,'HepatitisB':20,'HepatitisC':21,'HepatitisD':22,'Hepatitis:23,'Alcoholichepatitis':24,'Tuberculosis':25,’ (vertigo) Paroymsal Positional Vertigo':36,'Acne':37,'Urinary tract infection':38,'Psoriasis':39,'Impetigo':40}},inplace=True)

# print(df.head())

X= df[l1]

y = df[["prognosis"]]

np.ravel(y)

# print(y)

tr=pd.read\_csv("Testing.csv")

tr.replace({'prognosis':{'Fungal infection':0,'Allergy':1,'GERD':2,'Chronic cholestasis':3,'Drug Reaction':4,'Pepticulcerdiseae':5,'AIDS':6,'Diabetes':7,'Gastroenteritis':8,'BronchialAsthma':9,'Hypertension':10,'Migraine':11,'Cervicalspondylosis':12,'Paralysis(brainhemorrhage)':13,'Jaundice':14,'Malaria':15,'Chickenpox':16,'Dengue':17,'Typhoid':18,'hepatitisA':19,'Hepatitis B':20,'HepatitisC':21,'HepatitisD':22,'HepatitisE':23,'Alcoholichepatitis':24,'Tuberculosis':25,'CommonCold':26,'Pneumonia':27,'Dimorphichemmorhoids(piles)':28,'Heartattack':29,'Varicoseveins':30,'Hypothyroidism':31,'Hyperthyroidism':32,'Hypoglycemia':33,'Osteoarthristis':34,'Arthritis':35,'(vertigo) Paroymsal Positional Vertigo':36,'Acne':37,'Urinary tract infection':38,'Psoriasis':39,'Impetigo':40}},inplace=True)

X\_test= tr[l1]

y\_test = tr[["prognosis"]]

np.ravel(y\_test

def DecisionTree():

from sklearn import tree

clf3 = tree.DecisionTreeClassifier()

clf3 = clf3.fit(X,y)

from sklearn.metrics import accuracy\_score

y\_pred=clf3.predict(X\_test)

print(accuracy\_score(y\_test, y\_pred))

print(accuracy\_score(y\_test, y\_pred,normalize=False))

psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

for k in range(0,len(l1)):

# print (k,)

for z in psymptoms:

if(z==l1[k]):

l2[k]=1

inputtest = [l2]

predict = clf3.predict(inputtest)

predicted=predict[0]

h='no'

for a in range(0,len(disease)):

if(predicted == a):

h='yes'

break

if (h=='yes'):

t1.delete("1.0", END)

t1.insert(END, disease[a])

else:

t1.delete("1.0", END)

t1.insert(END, "Not Found")

def NaiveBayes():

from sklearn.naive\_bayes import GaussianNB

gnb = GaussianNB()

gnb=gnb.fit(X,np.ravel(y))

from sklearn.metrics import accuracy\_score

y\_pred=gnb.predict(X\_test)

print(accuracy\_score(y\_test, y\_pred))

print(accuracy\_score(y\_test, y\_pred,normalize=False))

psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

for k in range(0,len(l1)):

for z in psymptoms:

if(z==l1[k]):

l2[k]=1

inputtest = [l2]

predict = gnb.predict(inputtest)

predicted=predict[0]

h='no'

for a in range(0,len(disease)):

if(predicted == a):

h='yes'

break

if (h=='yes'):

t3.delete("1.0", END)

t3.insert(END, disease[a])

else:

t3.delete("1.0", END)

t3.insert(END, "Not Found”)

root = Tk()

root.configure()

Symptom1 = StringVar()

Symptom1.set(None)

Symptom2 = StringVar()

Symptom2.set(None)

Symptom3 = StringVar()

Symptom3.set(None)

Symptom4 = StringVar()

Symptom4.set(None)

Symptom5 = StringVar()

Symptom5.set(None)

Name = StringVar()

w2 = Label(root, justify=LEFT, text="Disease Predictor using Machine Learning", fg="white", bg="blue")

w2.config(font=("Elephant", 30))

w2.grid(row=1, column=0, columnspan=2, padx=100)

NameLb = Label(root, text="Name of the Patient", fg="yellow", bg="black")

NameLb.grid(row=6, column=0, pady=15, sticky=W)

S1Lb = Label(root, text="Symptom 1", fg="yellow", bg="black")

S1Lb.grid(row=7, column=0, pady=10, sticky=W)

S2Lb = Label(root, text="Symptom 2", fg="yellow", bg="black")

S2Lb.grid(row=8, column=0, pady=10, sticky=W)

S3Lb = Label(root, text="Symptom 3", fg="yellow", bg="black")

S3Lb.grid(row=9, column=0, pady=10, sticky=W)

S4Lb = Label(root, text="Symptom 4", fg="yellow", bg="black")

S4Lb.grid(row=10, column=0, pady=10, sticky=W)

S5Lb = Label(root, text="Symptom 5", fg="yellow", bg="black")

S5Lb.grid(row=11, column=0, pady=10, sticky=W)

lrLb = Label(root, text="DecisionTree", fg="white", bg="red")

lrLb.grid(row=15, column=0, pady=10,sticky=W)

ranfLb = Label(root, text="NaiveBayes", fg="white", bg="red")

ranfLb.grid(row=19, column=0, pady=10, sticky=W)

OPTIONS = sorted(l1)

NameEn = Entry(root, textvariable=Name)

NameEn.grid(row=6, column=1)

S1En = OptionMenu(root, Symptom1,\*OPTIONS)

S1En.grid(row=7, column=1)

S2En = OptionMenu(root, Symptom2,\*OPTIONS)

S2En.grid(row=8, column=1)

S3En = OptionMenu(root, Symptom3,\*OPTIONS)

S3En.grid(row=9, column=1)

S4En = OptionMenu(root, Symptom4,\*OPTIONS)

S4En.grid(row=10, column=1)

S5En = OptionMenu(root, Symptom5,\*OPTIONS)

S5En.grid(row=11, column=1)

dst = Button(root, text="DecisionTree", command=DecisionTree,bg="green",fg="yellow")

dst.grid(row=8, column=3,padx=10)

lr = Button(root, text="NaiveBayes", command=NaiveBayes,bg="green",fg="yellow")

lr.grid(row=10, column=3,padx=10)

t1 = Text(root, height=1, width=40,bg="orange",fg="black")

t1.grid(row=15, column=1, padx=10)

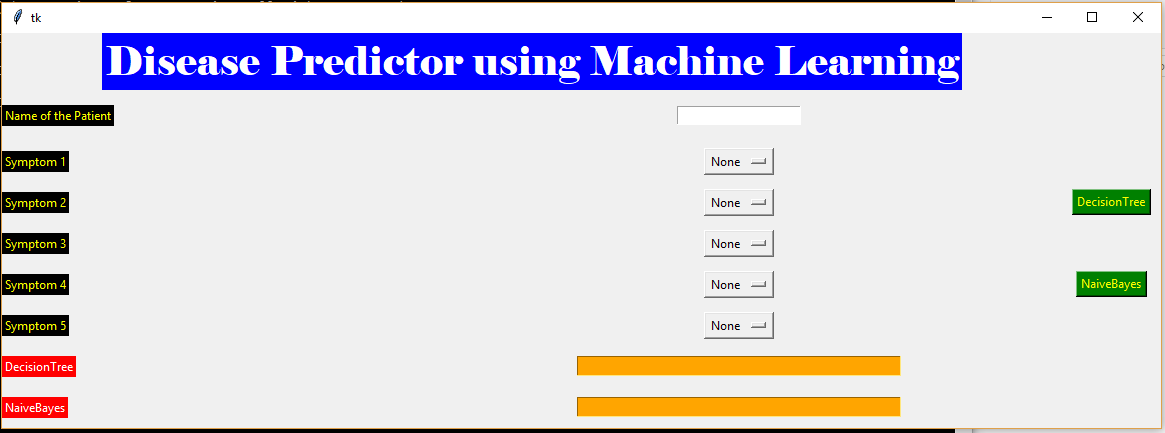
t3 = Text(root, height=1, width=40,bg="orange",fg="black")

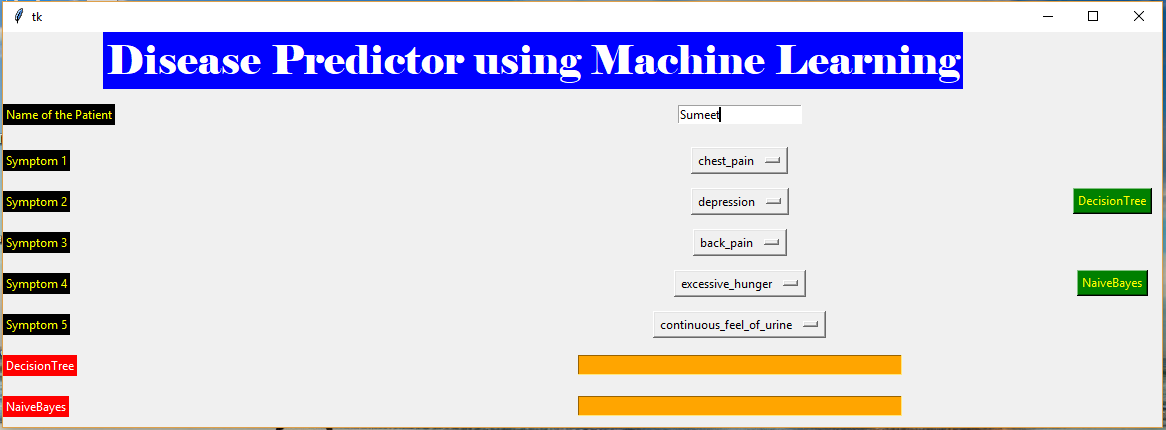
t3.grid(row=19, column=1 , padx=10)

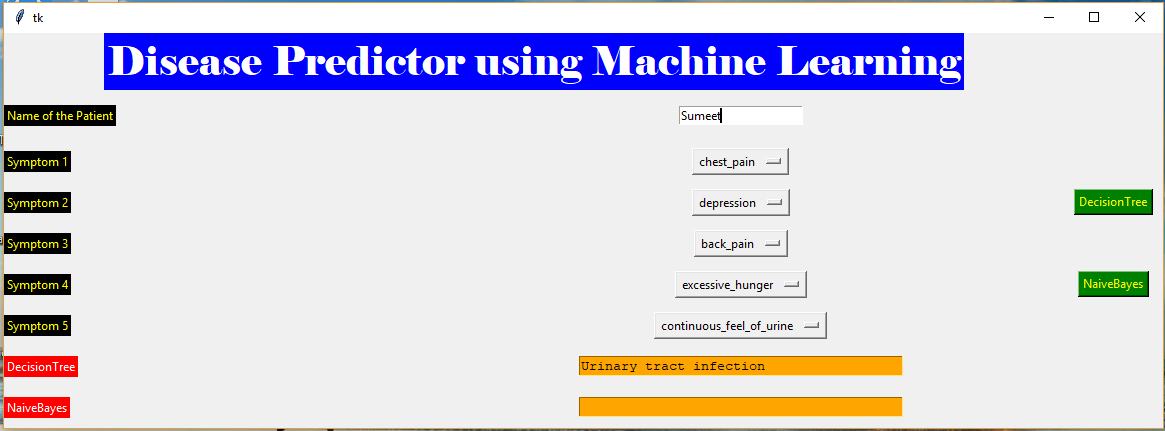
root.mainloop()

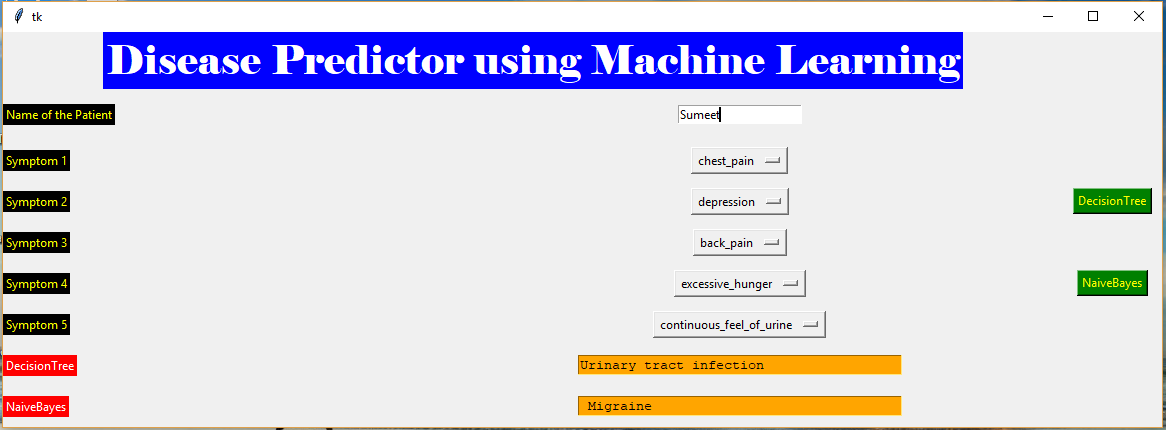
***RESULTS AND DISCUSSIONS:***

1. ***Result –***





******

******

***DISCUSSION-***

The five symptoms of the respective patient are inputted by the patient with its name .On the basis of those symptoms,by selecting the respective Machine Learning algorithm i.e. Decision Tree or Naïve Baye’s Classifier,the system predicts the disease of the respective patient.The disease predicted by both the algorithms may not be exact same but will be nearby to the exact disease as each algorithm runs according to its rules.

***CONCLUSIONS AND FUTURE SCOPE:***

1. ***Conclusion-***

We’ve successfully learned the concept of how machine learning algorithms are applied and implemented.

1. ***Future Scope –***

We can implement an application for such predictions which would help people to predict their disease on the basis of symptoms without having a late in increasing the seriousness of that disease.