REPORT

DISEASE PREDICTON USING MACHINE LEARING

NAME- Dev Chauhan
UNIVERSITY ROLL NO- 2013648
SECTION- DS
CLASS ROLLNO.-29

Problem Statement

AS THE WORLD IS CHANGING SO ARE THE NEEDS OF HELTHACRE DEPARTMENT TODAY THE CHALLENGES IS FOR MANY PEOPLE WHO ARE LOOKING ONLINE FOR HEALTH INFO. REGARDING DISEASE DIAGNOSIS AND DIFFERENT TREATMENTS IF A RECOMMENDED SYSTEM CAN BE MADE FOR THE DOCTORS AND MEDECINE WITH USING REVIEW MINING WILL SAVE A LOT OF TIME.

TOOLS AND LIBRABIBS

1. Pandas: It is a library available for python programming generally used for data analysis or while dealing with data.

2. Matplotlib: It is a library available for python programming used to visualize the data such as to plot the graphs, bar charts, etc.

3. NumPy: It is a library available for python programming used for dealing with matrices and arrays and to apply mathematical operations and functions on our dataset in the present in the form of an array.

4. Seaborn: Seaborn is a library for making statistical graphics in Python. It builds on top of Matplotlib and integrates closely with pandas data structures.

5. Math: This module provides access to the mathematical functions defined by the C standard.

6. TKINTER: this model is used for the Grapical user interface in python.

7. Standard scaler: Is used for standardize features by removing the mean and scaling to unit variance.

METHODOLOGY

These are the imported libraries that are utilized to use various tools that are available in that specific library. Tkinter is used to build a Graphical User Interface in Python.

L1 is the list made for various Symptoms which are generally showed up in people for various Diseases.

Disease is the list made for different Diseases which are for the most part appeared in different individuals.

First L2 is the vacant list made. At that point, equivalent to a number of diseases in list L1, L2 is appended in a number of zeroes.

There is a CSV document containing diseases and symptoms, named training.csv, which is utilized to prepare the model. Read_csv() function is utilized to store the information in the dataframe, named df. Utilizing replace() function, prognosis column that are the different diseases, it is replaced by the numbers from 0 to n-1, where n is the number of different diseases present in .csv record. Head() function is utilized to print the initial five rows of the preparation dataframe.

the output produced which contains the initial five rows of the dataframe df.

Putting the Symptoms in X and prognosis/diseases in y for training the model.

Output for the print(y) in whci different disease has values according to their symptoms.

To build the precision of the model, we utilized four distinctive algorithms which are as per the following

- Decision Tree algorithm
- Random Forest algorithm
- KNearestNeighbour algorithm
- Naive Bayes algorithm

Root=Tk() is used to for start working with the tkinter to build the gui. Definition of DecisionTree() function. “pred1” is used to store the predicted disease using decision tree algorithm

If user tries to run the gui without entering the name, then System will prompt the following message.

After filling the name, user have to fill five symptoms and out of which first two are compulsory. If user will not select atleast two symptoms, then following message will be prompt from the system

DecisionTreeClassifier() is used to train the model and predict the disease on testing dataset according to symptoms entered by the user. Final disease for decision tree is stored in a variable named “pred1”. Accuracy of predicting the disease is printed using accuracy_score and confusion matrix is created using confusion_matrix which are imported from sklearn.metrices. Definition of randomforest() function. “pred2” is used to store the predicted disease using random forest algorithm.

RandomForestClassifier() is used to train the model and predict the disease on testing dataset according to symptoms entered by the user. Final disease for random forest is stored in a variable named “pred2”. Accuracy of predicting the disease is calculated using accuracy_score and confusion matrix is created using confusion_matrix which are imported from sklearn.metrices.

Definition of KNN() function. “pred4” is used to store the predicted disease using kNearestNeighbour algorithmKNeighboursClassifier() is used to train the model and predict the disease on testing dataset according to symptoms entered by the user. Final disease for

kNearestNeighbour is stored in a variable named “pred4”. Accuracy of predicting the disease is calculated using accuracy_score and confusion matrix is created using confusion_matrix which are imported from sklearn.metrices

Definition of NaiveBayes() function. “pred3” is used to store the predicted disease using Naïve Bayes algorithm. GaussianNB() is used to train the model and predict the disease on testing dataset according to symptoms entered by the user. Final disease for Naïve Bayes is stored in a variable named “pred3”. Accuracy of predicting the disease is calculated using accuracy_score and confusion matrix is created using confusion_matrix which are imported from sklearn.metrices.

Graphical User Interface is build using tkinter library in Python. Root is used to start the GUI. It is configured with the background that is set to “Ivory”. GUI titlt is given as “Smart Disease Predictor System” using title() function in tkinter library. Resizable function is used to fix the size GUI.

Here, variables are defined like Name, Symptom1, Symptom2, etc and they initialised to “Select Here” using set() function in tkinter library

“Reset()” which is used to reset the GUI inputs which are given by the user. It is called when user click on the button “Reset Inputs” from the GUI.

“Exit()” which is used to come out from the GUI. It is called when user click on the button “Exit System” from the GUI

“W2” is the label created for showing the headings in the GUI using label() function from tkinter library. Two text are written under label w2 in row1 and row2

These are the labels for showing the Symptoms of the disease. It is created using label() function from tkinter library. Its features are configured using config() function and their grid is set by using grid() function from tkinter library.

NameEn is the entry box created for getting the name of the patient using Entry() function in tknirter library. S1, S2, S3, S4, S5 are the option menu used to get symtoms from the user which is created using Optionmenu in tkinter library. *OPTIONS is the list of unique symtoms.

Root.mainloop() is used to call all the functions.