Report

Diseases Prediction

DISEASE PREDICTION BASED ON SYMPTOMS

Machine Learning:Supervised learning problem

Students: Bassole Cedric-Francois

University:ISET BIZERTE

Class:Master in Robotics and Artificial Intelligence

Course: Machine Learning

2023 Mai 28

Abstract

Any health-related issue must be accurately and promptly analyzed if it is to be prevented and treated. In the event of a critical illness, the conventional method of diagnosis might not be sufficient. A more accurate diagnosis than the traditional approach can be achieved by creating a medical diagnosis system based on machine learning algorithms for prediction of any disease. With the use of ML algorithms, we have created a system for disease prediction. More than 230 diseases,their symptoms severity,their symptoms description and their symptoms precautions were present in 4 dataset that was processed. The diagnosis system produces the condition that a person may be suffering from,based on several symptoms.

Contents

1. Motivation

The economy and the welfare of humanity depend on a functional healthcare system. There has been a significant amount of change between the world we live in today and the one we did a few decades ago. Everything has become more disorganized and ugly. In this case, medical professionals are risking their own lives in order to save as many lives as they possibly can.

Board-certified medical professionals known as virtual doctors prefer to do phone and video consultations over in-person consultations when they can, however this is not always possible in an emergency. Machines are considered to be superior to humans in the absence of human error because they can do tasks more quickly while keeping a constant degree of precision. Without involving a person, a disease predictor, also referred to as a virtual doctor, can correctly forecast a patient's illness. In severe cases, like COVID-19 and EBOLA, a disease predictor can save a person's life by identifying their health without the need for physical contact. There are virtual doctors available now, but they lack the ability to offer the necessary level of precision.

Doctors may make mistakes when diagnosing a patient's disease, but disease prediction systems with machine learning algorithms can help produce accurate results in these situations. For this project, we employed a mix of approaches , algorithms, and technologies to develop a system that can forecast a patient's status based on hospital data and machine learning methods based on the Python programming language. Following that, the data is categorized using the K-NN algorithm.

2. Approach

2.1 Dataset

This inquiry used data from Kaggle.It's A dataset to provide the students a source to create a healthcare related system There are columns containing diseases, their symptoms , precautions to be taken, and their weights. This dataset can be easily cleaned by using file handling in any language. The user only needs to understand how rows and coloumns are arranged.

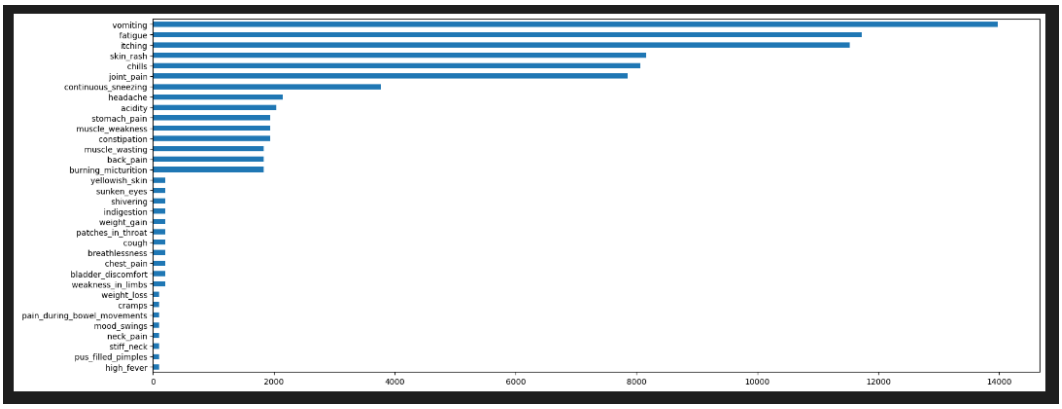

Figure 1: Image of desease distribution in the dataset

2.2 Proposed System

In the suggested strategy, we use Machine Learning techniques and a web interfaceto precisely forecast the ailment that the patient has been suffering from. When past healthcare records are used as a dataset, the results are more accurate. To train the model and predict user diseases based on the symptoms they enter, we use machine learning algorithms.

Advantages of Proposed System

- **First and foremost, seeing a doctor for modest treatment is unnecessary.**
- **When compared to past treatments, you'll get more precise results**
- **Only a few risk variables are at play**

3. Model and algorithm

To construct a disease prediction based on symptoms, we applied 2 machine learning algorithms: Random Forest, KNN. We can get an accurate forecast for our model using these tactics. The Prognosis of the Illness Currently, the effort is in full swing. Machine Learning is being used to diagnose and prevent disease in its infancy. As we all know, humanity has become so engrossed in the competitive environment of economic advancement that it has lost sight of its own well-being. Studies show that 40 percent of people ignore small symptoms, which might lead to more serious problems in the future. The project's interface is also built with Flutter. The user must first enter their name, then select symptoms cases; alternatively, the user must enter all symptoms, after which the system will return an exact result. Two machine learning approaches were used to create this forecast: Random Forest, KNN. When the user enters all of the symptoms and simply presses the Predict Button, the result is computed ; similarly, we've utilized four ways to provide a more thorough perspective of the data, and the user must be satisfied with the anticipated conclusion.

3.1 Activation function

Random Forest

Random Forest, a well-known machine learning algorithm, employs the supervised learning method. In machine learning, it can be utilized for both classification and regression issues. It is based on ensemble learning, which is a method for solving a complicated problem by merging numerous classifiers and improving the model's performance.

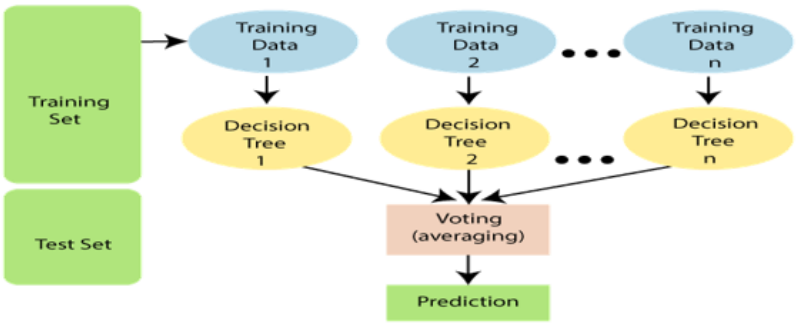

Figure 2: sigmoid and ReLu

KNN

One of the most fundamental Machine Learning algorithms is the K-Nearest Neighbour approach. It is based on the method of Supervised Learning. Because K-NN considers the new case/data and previous cases to be comparable, the new case is assigned to the category that is the most similar to the previous categories.
The K-NN method keeps track of all available data and categorizes new data points based on how similar they are to existing data. As fresh data arrives, the K-NN algorithm can quickly filter it into the appropriate suite category. Although this method can be used for both regression and Classification, classification is the most popular use.

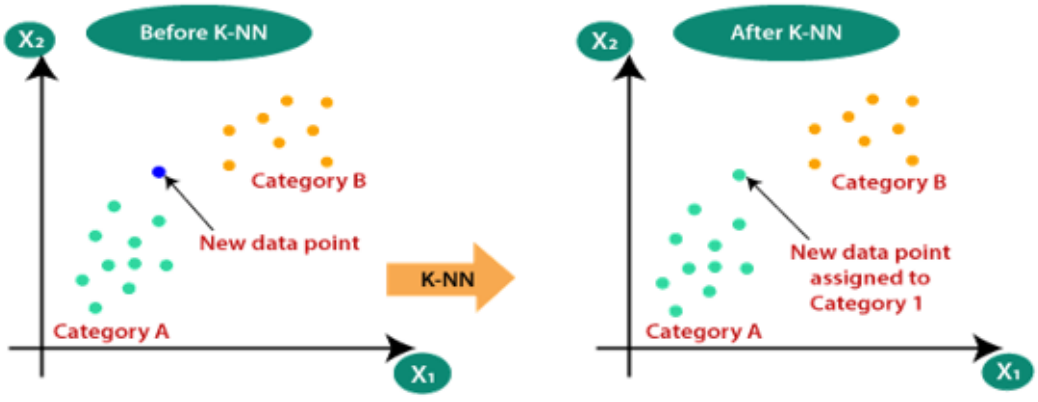

Figure 3: Neural Network

4. Experiment and Results

4.1 Experimentation

To conduct all of the experiments in the Jupyter notebook, we used the python3 programming language on one hand and for testing purpose we used Azur Machine learning with an embedded python script to deploy a webservice Later

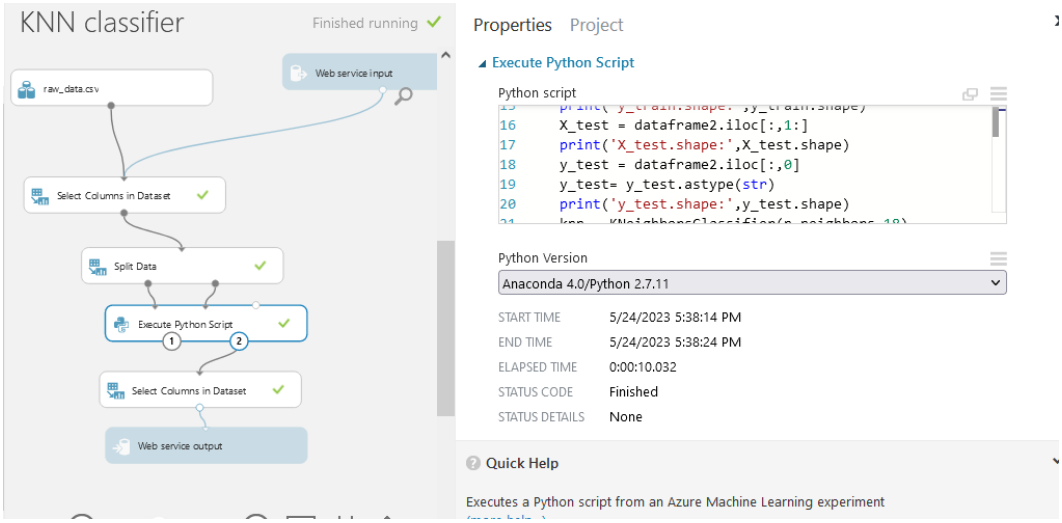

Figure 4: Azure Machine learning

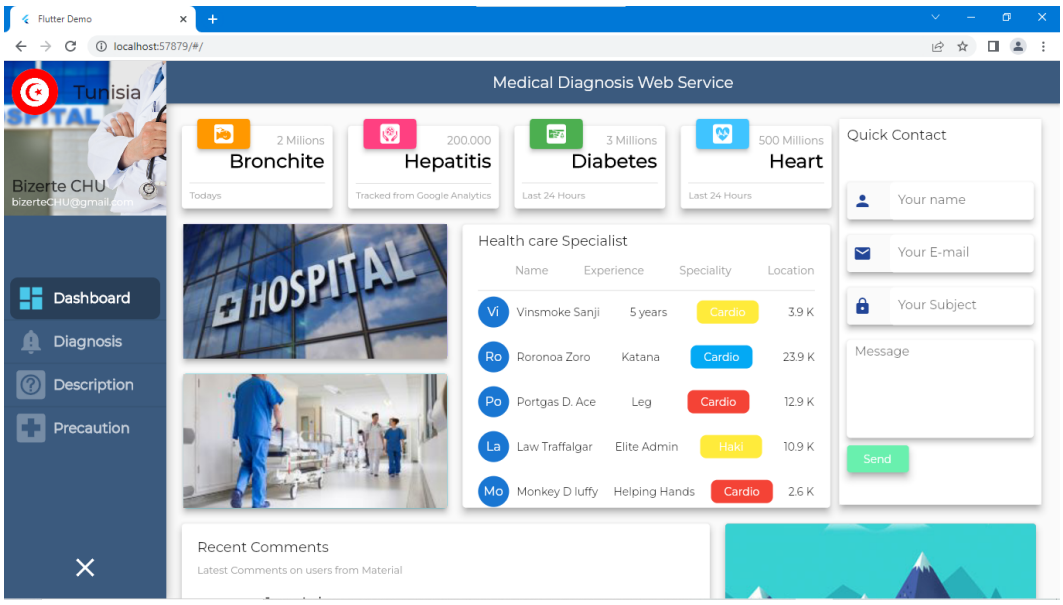

Figure 5: Web interaface build with Flutter

4.2 Metrics for Assessment

Metrics for Assessment

We get accurate disease prediction because we supply symptoms as input to the system

```
print('F1-score% =', f1_score(y_test, result, average='macro')*100, '|', 'Accuracy% =', accuracy_score(y_test, result)*100)
```

```
F1-score% = 95.32311193280204 | Accuracy% = 95.25745257452574
```

Figure 6: Confusion Matrix

4.3 Disease Prediction Dataset

Disease Prediction Dataset

A CSV data file from New York-Presbyterian Hospital was provided by the University of Columbia. The Deseases Symptoms data file has 4920 rows and 18 columns while the deseases severity have 133 rows and 2 columns. Itching, skin rash, shivering, chills, joint stiffness, and other symptoms are some of the most prevalent attributes. After the preprocessing step There is a total of 18 columns in the dataset out of which 17 columns represent the symptoms and the last column is the prognosis.

```
data_sevrity.head(10)
```

	Symptom	weight
0	itching	1
1	skin_rash	3
2	nodal_skin_eruptions	4
3	continuous_sneezing	4
4	shivering	5
5	chills	3
6	joint_pain	3
7	stomach_pain	5
8	acidity	3
9	ulcers_on_tongue	4

Figure 7: Severity Dataset

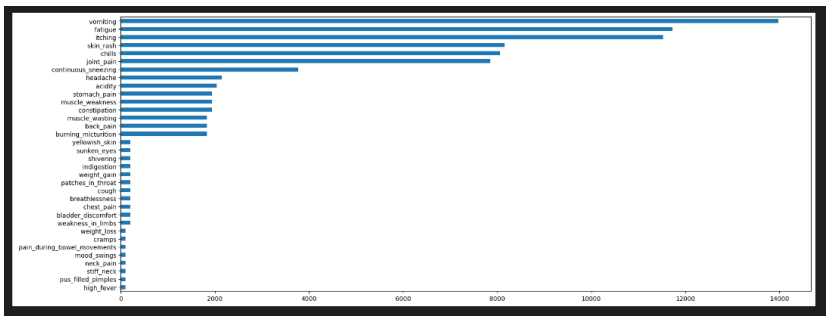

Figure 8: Reaprtition Symptoms in dataset

4.4 Data Preprocessing

This step will remove any punctuation,delete null,replace symptoms by their weigth

```
def remove_space_between_word(dataset):
    for col in dataset.columns:
        for i in range(len(dataset[col])):
            if (type(dataset[col][i]) == str ):
                dataset[col][i] = dataset[col][i].strip()
                dataset[col][i] = dataset[col][i].replace(" ", "_")
    return data
```

```
new_df = remove_space_between_word(data)
new_df.head()
```

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_
0	Fungal_infection	itching	skin_rash	nodal_skin_eruptions	dischromic__patches	NaN	NaN
1	Fungal_infection	skin_rash	nodal_skin_eruptions	dischromic__patches	NaN	NaN	NaN
2	Fungal_infection	itching	nodal_skin_eruptions	dischromic__patches	NaN	NaN	NaN
3	Fungal_infection	itching	skin_rash	dischromic__patches	NaN	NaN	NaN
4	Fungal_infection	itching	skin_rash	nodal_skin_eruptions	NaN	NaN	NaN

Figure 9: Preprocessing

4.5 Training/Results

The system will compare the user's symptoms to the dataset as they are entered, the dataset is made up of binary 0s and 1s, and once the model has assessed all of the user's symptoms, it will accurately forecast the disease associated with that manifestation.

```
x_train, x_test, y_train, y_test = train_test_split(df_data, label, shuffle=True, train_size = 0.70)
knn = KNeighborsClassifier(n_neighbors=18)
randomFC = RandomForestClassifier()
knn.fit(x_train, y_train)
result = knn.predict(x_test)

print('F1-score% =', f1_score(y_test, result, average='macro')*100, '|', 'Accuracy% =', accuracy_score(y_test, result)*100)
```

```
F1-score% = 95.32311193280204 | Accuracy% = 95.25745257452574
```

Figure 10: Training Results

4.6 Prediction/ Outputs

The system will compare the user's symptoms to the dataset as they are entered, the dataset is made up of binary 0s and 1s, and once the model has assessed all of the user's symptoms, it will accurately forecast the disease associated with that manifestation.

	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7	Symptom_8	Symptom_9	Symptom_10	Symptom_11
0	1	2	3	4	5	6	1	2	3	4	5

```
[23] output = knn.predict(qw)

[24] output[0]

'Tuberculosis'
```

Figure 11: Exemple of Prediction

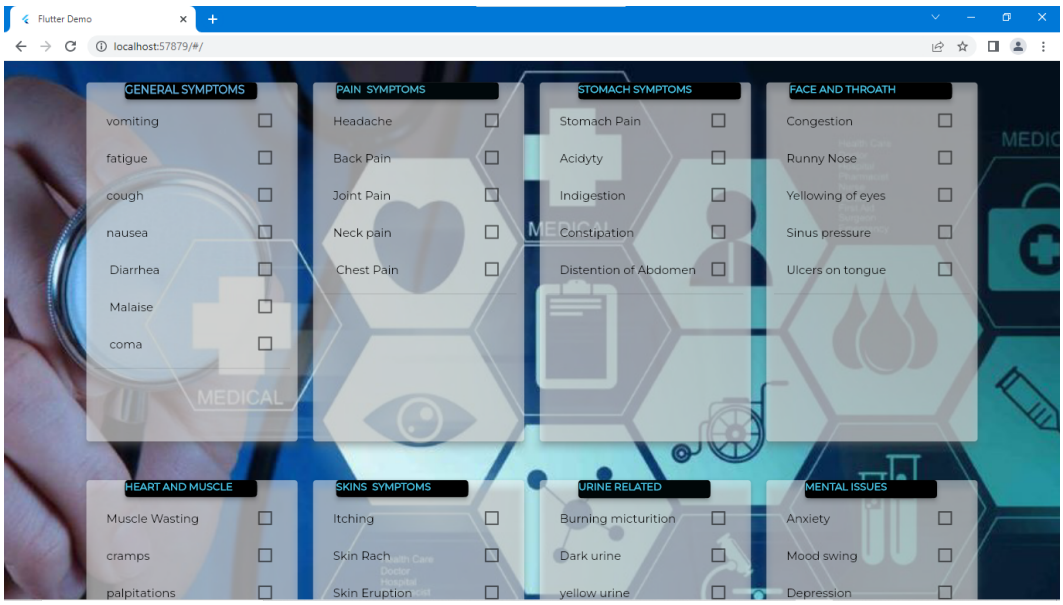

Figure 12: Web interface diagnosis page

5. Conclusion

Last but not least, I'd like to stress how crucial this project—disease prediction using machine learning—is to everyone's daily lives, but especially to those in the healthcare sector, who use these systems frequently to forecast patients' diseases based on their general characteristics and symptoms. Since the health sector now plays such a significant role in treating patients' illnesses, it is frequently quite beneficial for the sector to inform the user. It is also helpful for the user if he or she does not want to visit a hospital or other clinics because the user can learn about the disease they are experiencing by simply entering the symptoms and any other pertinent information, and the sector can also profit from this system. If the healthcare sector adopts this idea, doctors' workloads will be reduced and they will be better able to predict a patient's illness.

A method known as disease prediction allows doctors to anticipate the development of a variety of common diseases that, if untreated or ignored, can cause death and a host of other issues for the patient and their family.