

Disease Identifier by Symptoms Using Machine Learning

Prof Bhavesh Shah , Jai singh and Akansha Shelar

Suman Ramesh Tulsiani Technical Campus and Faculty of Engineering

Project Description

Health data wants are dynamical the information seeking behaviour and should be observed around the globe. Challenges long-faced by many of us are wanting on-line for health information relating to diseases, diagnoses and totally different treatments. If a recommendation system can be created for doctors and medication whereas mistreatment review mining can save lots of a while. In this system like these, the user faces several issues in understanding the core medical vocabulary as the users are laymen. User is confused as a result of Associate in Nursing oversize quantity of medical data on different mediums are offered. The idea behind recommender system is to adapt to have an effect on the special needs of the health issues of a user.

Introduction

With the increase in variety of patient and malady once a year medical system is full and with time became expensive in several countries. Most of the malady involves a consultation with doctors to induce treated. With spare information prediction of malady by Associate in Nursing formula will be terribly straightforward and cheap. Prediction of malady by viewing the symptoms is Associate in Nursing integral a part of treatment. In

our project we've got tried accurately predict a malady by viewing the symptoms of the patient.

We have used four totally different algorithms for this purpose Associate in Nursing gained an accuracy of 2-95%. Such a system will have a awfully massive potential in medical treatment of the longer term. we've got conjointly designed an interactive interface to facilitate interaction with the system. we've got conjointly tried to point out and visualised the results of our study and this project. Database assortment Dataset for this project was collected from a study of university of Columbia performed at New York Protestant Hospital throughout 2004. Link of dataset is given below.

<http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html>

In this project customary libraries for info analysis and model creation square measure used. the subsequent

are the libraries utilized in this project.

1. tkinter: It's a typical graphical user interface library of python. Python once combined with tkinter provides fast and straightforward thanks to produce graphical user interface. It provides powerful object-oriented tool for making graphical user interface.

It provides numerous widgets to make graphical user interface a number of the distinguished ones being:

- 1) Button
- 2) Canvas
- 3) Label
- 4) Entry
- 5) Check Button
- 6) List box
- 7) Message
- 8) Text
- 9) Messagebox

Some of these were utilized in this project to make our graphical user interface particularly messagebox, button, label, Option Menu, text and title. victimization tkinter we have a tendency to were ready to produce associate interactive graphical user interface for our model.

2. Numpy: Numpy is core library of scientific computing in python. It provides powerful tools to deal with numerous multi-dimensional arrays in python. it's a general purpose array process package. Numpy's main purpose is to alter two-dimensional homogeneous array. it's tools go from array creation to its handling. It makes it easier to make a n dimensional array simply by victimization np.zeros() or handle its contents victimization numerous different strategies like replace, arrange, random, save, load it conjointly helps I array process victimization strategies like total, mean, std, max, min, all, etc Array created with numpy conjointly behave otherwise then arrays created ordinarily after they square measure operated upon victimization operators like +,-,*,/.

All the higher than qualities and services offered by numpy array makes it extremely appropriate for our purpose of handling information. information manipulation occurring in arrays whereas playing numerous operations got to offer the specified results whereas predicting outputs require such high operational capabilities.

3. pandas : it's the foremost well-liked python library used for information analysis. It provides extremely optimized performance with back-end ASCII text file strictly written in C or python.

Data in python may be analysed with two ways that

- Series
- Dataframes

Series is one dimensional array outlined in pandas accustomed store any information kind.

Dataframes square measure two-dimensional organization utilized in python to store information consisting of rows and columns. Pandas dataframe is employed extensively during this project to use datasets needed for coaching and testing the algorithms. Dataframes makes it easier to figure with attributes and results. Several of

its intrinsic functions like replace were utilized in our project for information manipulation and preprocessing.

4. sklearn: Sklearn is associate open supply python library with implements a large vary of machine- learning, pre-processing, cross-validation and visualisation algorithms. It options numerous easy and economical tools for data processing and processing. It options numerous classification, regression and clump rule like support vector machine, random forest classifier, decision tree, Gaussian naïve-Bayes, KNN to call a number of.

In this project we've got used sklearn to induce advantage of intrinsic classification algorithms like decision tree, random forest classifier, KNN and naïve Thomas Bayes. we've got conjointly used intrinsic cross validation and visualisation options like classification report, confusion matrix and accuracy score.

Models

There square measure four completely different reasonably models gift in our project to predict the illness these square measure

- call tree
- Random forest tree
- Gaussian Naïve Thomas Bayes
- KNN

Decision tree is assessed as a awfully effective and versatile classification technique. it's utilized in pattern recognition and classification for image. it's used for classification in terribly complicated problems condensation to its high ability. it's conjointly capable of participating issues of upper dimensionality. It primarily consists of 3 elements root, nodes and leaf. Roots consists of attribute that has most impact on the end result, leaf tests for price of sure attribute and leaf offers out the output of tree.

Decision tree is that the initial prediction technique we've got utilized in our project. It offers United States associate accuracy of ~95%.

Random Forest rule may be a supervised learning rule used for each classification and regression. This rule works on four basic steps –

1. It chooses random information samples from dataset.
2. It constructs call trees for each sample dataset chosen.
3. At this step each foretold result are compiled and voted on.
4. eventually most voted prediction are selected and be bestowed as results of classification. In this project we've got used random forest classifier with a hundred random samples and therefore the result given is ~95% accuracy. K Nearest Neighbour may be a supervised learning rule. it's a basic nonetheless essential rule. It finds intensive use in pattern finding and data processing.

It works by finding a pattern in information that links information to results and it improves upon the patter recognition with each iteration. We have used K Nearest Neighbour to classify our dataset and achieved ~92% accuracy. Naïve Thomas Bayes rule may be a family of algorithms supported naïve Thomas Bayes theorem. They share a common principle that's each try of prediction is freelance of every different. It conjointly makes associate assumption that options create associate freelance and equal contribution to the prediction.

In our project we've got used naïve Thomas Bayes rule to achieve a ~95% correct prediction.

Disease Predictor using Machine Learning
Contributors: Akanksha And Jai

Name of the Patient *	<input type="text" value="abd"/>	Prediction 1
Symptom 1 *	<input type="text" value="constipation"/>	Prediction 2
Symptom 2 *	<input type="text" value="cramps"/>	Prediction 3
Symptom 3	<input type="text" value="Select Here"/>	Prediction 4
Symptom 4	<input type="text" value="Select Here"/>	Reset Inputs
Symptom 5	<input type="text" value="Select Here"/>	Exit System

DecisionTree	Dimorphic hemmorhoids(piles)
RandomForest	Drug Reaction
NaiveBayes	Drug Reaction
kNearestNeighbour	Fungal infection

Disease Predictor using Machine Learning
Contributors: Akanksha And Jai

Name of the Patient *

Symptom 1 *

Symptom 2 *

Symptom 3

Symptom 4

Symptom 5

DecisionTree

RandomForest

NaiveBayes

kNearestNeighbour

Dimorphic hemmorhoids(piles)

Paralysis (brain hemorrhage)

Drug Reaction

Allergy

Prediction 1

Prediction 2

Prediction 3

Prediction 4

Reset Inputs

Exit System

Conclusions

We embarked on to make a system which might predict illness on the idea of symptoms given thereto. Such a system will decrease the push at OPDs of hospitals and scale back the employment on medical staff. we have a tendency to were booming in making such a system and use four completely different rule to try to to thus. On an average we have a tendency to achieved accuracy of ~94%. Such a system may be mostly reliable to try to to the work. Creating this method we have a tendency to conjointly else the way to store the information entered by the user within the info which can be utilized in future to assist in making higher version of such system. Our system conjointly has an easy to use interface. It conjointly has numerous visual illustration of information collected and results achieved.

REFERENCES

[1]Dahiwade, D., Patel, G., & Meshram, E. (2019). Designing Disease Prediction Model Using Machine Learning Approach. 2019 3rd International Conference on Computing Methodologies and Communication.

[2] Kohli, P. S., & Arora, S. (2018). Application of Machine Learning in Disease Prediction. 2018 4th International Conference on Computing Communication and Automation.

[3] Huang, F., Wang, S., & Chan, C.-C. (2012). Predicting disease by using data processing based on healthcare data system. 2012 IEEE International Conference on Granular Computing.

[4] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease Prediction by Machine Learning Over Big Data from Healthcare Communities.

[5] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). heart condition prediction using lazy associative



classification. 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing

[6] Bhavesh shah, Akanksha shelar & Jai Singh (2021). Disease Identifier By Symptoms Using Machine Learning.