Disease Prediction using Machine Learning

Sakshi Awankar[1], Aakanksha Darvankar[2], Vaishnavi Dayma[3], Prof. V. N. Bhonge[4]

Students, Department of Electronics and Telecommunication Engineering[1,2,3]
Guide, Department of Electronics and Telecommunication Engineering[4]
Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India

Abstract: *Evolution of modern technologies like data science and machine learning has opened the path for healthcare communities and medical institutions. The earth is going to a patch of technology where demand of intelligence and accuracy is increasing behind it. Today's people are likely addicted to internet but they are not concerned about their physical health. Most of the people are prone to diseases, due to lack of healthy food, proper sleep and daily exercise. People ignore the small problem and don't visit the hospital which turn into serious disease with time. So it is very crucial to know if we are suffering from a disease, at an early stage rather than discovering it at later stage. Hence disease prediction system plays an important role as it predicts the diseases based on symptoms. This disease prediction system uses Machine Learning algorithm named Naive Bayes Theorem.*

Keywords: Machine Learning, Disease Prediction, Naïve Bayes algorithm, Supervised Learning.

I. INTRODUCTION

With the advancement in technology, Machine Learning is becoming more popular and commonly used technology by industry experts for solving problems faced in real life. Machine Learning is the scientific study of algorithms and statistical models that computer use to perform a specific task without using explicit instructions, relying on patterns and inference instead. Machine Learning is also used by the healthcare industry to bring advancement in their techniques so that they can provide better services. Machine Learning is a subset of AI that is mainly deal with the study of algorithms which improve with the use of data and experience. Machine Learning has two phases i.e. Training and Testing. Machine Learning provides an efficient platform in medical field to solve various healthcare issues at a much faster rate. There are two kinds of Machine Learning – Supervised Learning and Unsupervised Learning. In supervised learning we frame a model with the help of data that is well labelled. On the other hand, unsupervised learning model learn from unlabeled data. In this, supervised Machine Learning concept is used for predicting the diseases.

II. PROPOSED MODEL

We are proposing such a system that can predict the possible diseases based on symptoms so it can be cured at the early stage. It saves time that is required to do the complete diagnosis of the patient and based on the suggestions provided by the system, we can only get the patient diagnosed for those disease that are required. In this, we are using machine learning algorithm that is Naive Bayes algorithm that try to accurately predict possible desease. The main purpose of this study is to help the non-technical person and freshman doctors to make a correct opinion about the diseases.

III. METHODOLOGY

3.1 Libraries Used

In this project standard libraries for database analysis and model creation are used. The following are the libraries used in this project.

A. tkinter

It is a standard GUI library of python. Python when combined with tkinter provides fast and easy way to create GUI. It provides powerful object-oriented tool for creating GUI. It provides various widgets to create GUI some of the prominent ones being:

- Button
- Canvas

- Label
- Check Butt
- List box
- Message
- Text

Some of these were used in this project to create our GUI namely messagebox, button, label, Option Menu, text and title. Using tkinter we were able to create an interactive GUI for our model.

B. Numpy

Numpy is core library of scientific computing in python. It provides powerful tools to deal with various multi-dimensional arrays in python. It is a general purpose array processing package. Numpy's main purpose is to deal with multidimensional homogeneous array. It has tools ranging from array creation to its handling. It makes it easier to create a n dimensional array just by using np.zeros() or handle its contents using various other methods such as replace, arrange, random, save, load it also helps I array processing using methods like sum, mean, std, max, min, all, etc Array created with numpy also behave differently then arrays created normally when they are operated upon using operators such as +,-,*,/.All the above qualities and services offered by numpy array makes it highly suitable for our purpose of handling data. Data manipulation occurring in arrays while performing various operations need to give the desired results while predicting outputs require such high operational capabilities.

C. pandas

It is the most popular python library used for data analysis. It provides highly optimized performance with back-end source code purely written in C or python. Pandas dataframe is used extensively in this project to use datasets required for training and testing the algorithms. Dataframes makes it easier to work with attributes and results. Several of its inbuilt functions such as replace were used in our project for data manipulation and preprocessing.

D. sklearn

Sklearn is an open source python library with implements a huge range of machine learning, pre-processing, cross-validation and visualization algorithms. It features various simple and efficient tools for data mining and data processing. It features various classification, regression and clustering algorithm such as support vector machine, random forest classifier, decision tree, gaussian naïve-Bayes, KNN to name a few. In this project we have used sklearn to get advantage of inbuilt classification algorithms like naïve Bayes.

3.2 Model
A. Naive Bayes Algorithm

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier. The Naive Bayes algorithmis called naive because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features. It is one of the most basic text classification techniques with various applications in email spam detection, personal email sorting, document categorization, language detection and sentiment detection. In spite of the naïve design and oversimplified assumptions that this technique uses, Naive Bayes performs well in many complex real-world problems. The basis of Naive Bayes algorithm is Bayes' theorem or alternatively known as Bayes' rule or Bayes' law. It gives us a method to calculate the conditional probability, i.e. the probability of an event based on previous knowledge available on the events. More formally, Bayes' Theorem is stated as the following equation:

$$P(A|B) = [P(B|A) \cdot P(A)] / P(B)$$

Another popular variation of Naïve Bayes is Multinomial Naïve Bayes algorithm. This algorithm estimates the conditional probability of a particular word/term/token given a class as the relative frequency of term in documents belonging to classand is given by the equation:

$$P(t|c) = T_{ct} / \sum t' \, \epsilon \, v \, T_{ct'}$$

Thus this variation takes into account the number of occurrences of term t in training documents from class c, including multiple occurrences. Usually Multinomial Naive Bayes is used when the multiple occurrences of the words matter a lot in the classification problem. Such an example is when we try to perform Topic Classification.
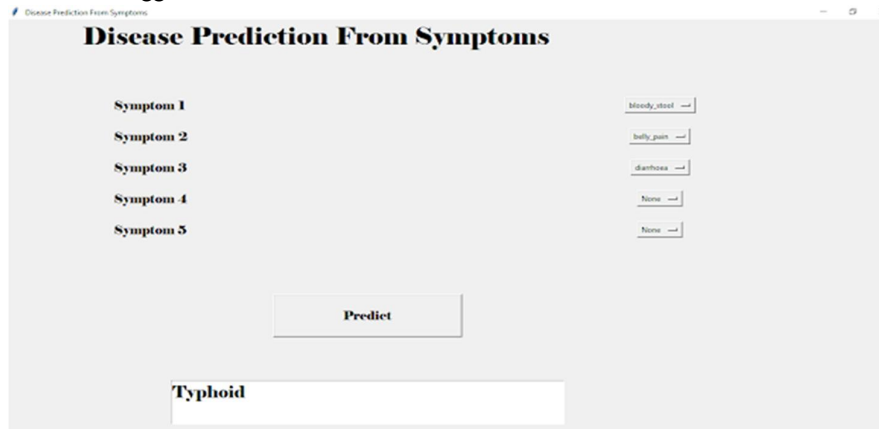
IV. SYSTEM ARCHITECTURE
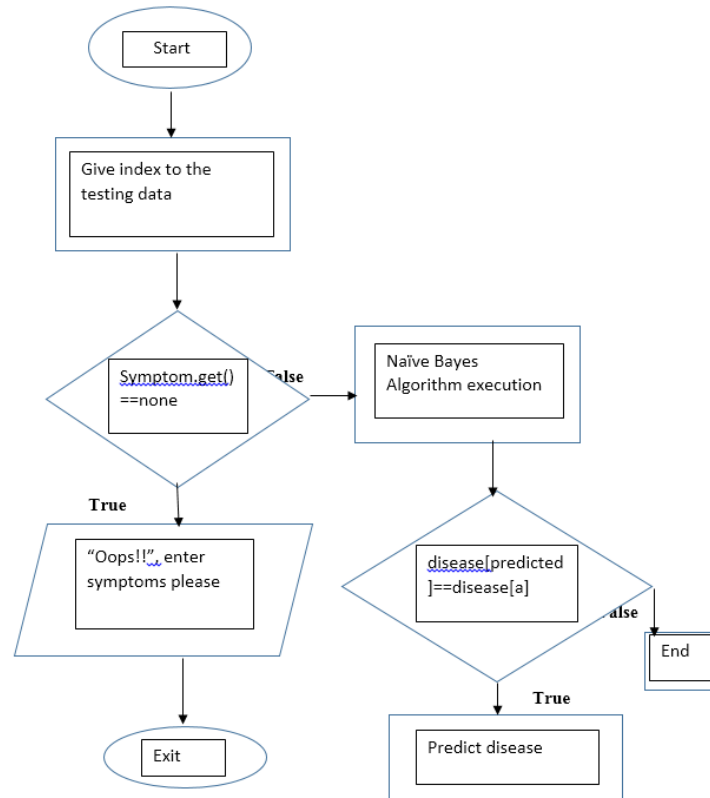
4.1 Steps to Develop the Project

1. Analyzing the problem statement & requirements Analyze the problem in terms of what we want to predict and what kind of observation data we have to make those predictions. Predictions are generally a label or a target answer; it may be a yes/no label (binary classification) or a category (multiclass classification) or a real number (regression).

2. Collect and clean the dataIdentify what kind of historical data we have for prediction modeling, the next step is to collect the data from datasets or from any other data sources.

3. Prepare data for ML application Transform the data in the form that the Machine Learning system can understand.

4. Prepare the Graphical User Interface (GUI) of the model Graphical User Interface (GUI) is designed for taking input and displaying output. There are 5 input text boxes which consist of dropdown menu of symptoms and the user can select those one by one. Python Tkinter package is used for designing the GUI. On pressing the 'Result' button, the disease is predicted in the output field. Also, the drugs are described in the specified field.

5. Train the model Before training the model, it is essential to split the data into training and evaluation sets, as we need to monitor how well a model generalizes to unseen data. Now, the algorithm will learn the pattern and mapping between the feature and the label.

6. Evaluate and improve model accuracy Accuracy is a measure to know how well or bad a model is doing on an unseen validation set. Based on the current learning, evaluate the model on validation sets.

7. Test the model Test the model on unknown data. After the system starts working properly, the model is complete.

Impact Factor: 6.252

4.2 Result

In the proposed system, a disease prediction model is built using a Machine Learning algorithm that is Naive Bayes Algorithm. Based on the symptoms that are input by the user, the disease is predicted and the drug that is most commonly prescribed by the doctor is suggested.

V. CONCLUSION

This project aims to predict the disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as input and produces output i.e. predict disease. In conclusion, for disease risk modelling, the accuracy of risk prediction depends on the diversity feature of the hospital data

REFERENCES

[1]. 2019 "Designing Disease Prediction Model Using Machine Learning Approach" , Dhiraj Dahiwade, Prof. Gajanan Patle, Prof. Ektaa Meshram.

[2]. 2018 "Application of Machine Learning in Disease Prediction" , Pahulpreet Singh Kohli, Shriya Arora.

[3]. 2020 "Disease Prediction using Machine Learning Algorithms" , Sneha Grampurohit , Chetan Sagarnal.

[4]. 2022 "Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models", C. Kavitha, Vinodhini Mani, S. R. Srividhya, Osamah Ibrahim Khalaf and Carlos Andrés Tavera Romero

[5]. 2020 "Disease Prediction Using Machine Learning", PRIYANKA J. PANCHAL, SHAEEZAH A. MHASKAR, TEJAL S. ZIMAN

[6]. 2022 "Identification and Prediction of Chronic Diseases Using Machine Learning Approach", Rayan Alanazi, Mohamed Elhoseny