**Abstract.** In this era of IT, technology has revolutionized the health domain to a great extent. This project aims to design a diagnostic model for various dis- eases relying on their symptoms. This System has used data mining techniques such as classification in order to achieve such a model. Datasets consisting of voluminous data about patient diseases are gathered, refined and classified and were used for training the intelligent agent. Here, the Naive Bayes Algorithm is used for classification purpose. Naïve Bayes Classifier calculates the probability of the disease. Based on the result, the patient can contact the doctor accordingly for further treatment. It's an exemplar where technology and health knowledge are sewn into a thread perfectly with a desire to achieve "prediction is better than cure".

**Keywords:** Naive bayes, medical data, classification, data mining

**1 Introduction**

Nowadays, the use of the internet has been stimulating curiosity among people and, be it of any kind, they are trying to find a solution to their problems through the inter- net only. It is a matter of fact that people have much easier access to the internet than hospitals and doctors. So, with the help of this system, a user can consult a doctor by sitting at their home itself. There will not be any fuss of visiting a clinic or hospitals and making your health condition worse. This Disease Prediction system is a web- based application that predicts the most probable disease of the user in accordance with the given symptoms by the help of the data-sets collected from different health- related sites. It often happens that someone nearer or dearer to you may need a doc- tor’s help immediately for some serious reasons but the doctor isn’t available for con- saltation for some prior commitments or other obvious reasons. That is when the role of this automated program comes into play. This Disease Prediction system can be used for urgent guidance on their illness according to the details and symptoms they will feed to the web-based application. Here, we use some intelligent data processing techniques to get the most accurate disease that would be related to the patient’s de- tails. And then based on the results, the patient can contact the respective disease specialist for any further treatments. This system can be used for a free consultation regarding any illness.



**Proposed system**

In the proposed system the diseases are predicted automatically by the system us- ing our model which is trained on the medical dataset. This system also shows the confidence score of the prediction. After getting the anticipated disease, the system will suggest doctors associated with that disease and therefore the patient can consult to the doctor online. The proposed system acts as a decision support system and will prove to be an aid for the physicians with the diagnosis.

**Fig.**



The disease prediction system has 3 users such as doctor, patient and admin. Each user of the system is authenticated by the system. The system allows the patient to give symptoms and according to those symptoms the system will predict a disease with a accuracy score. Then it suggests doctors for online consultation. Then the patient can consult a doctor anytime at his convenience.

**4. Algorithm implementation and evaluation:**

**4.1 Algorithm used**

Naive Bayes - There is a wide range of major algorithms for predicting various diseases with guessable symptoms in the field of Supervised Machine Learning and its different models such as Decision tree, Naive Bayes and Random Forest. Naïve Bayes has three different models i.e. Gaussian, Multinomial and Bernoulli NaïveBayes. Each model has its own accuracy to predict the result of diseases and the application of data fitting is mostly the same in all the three models. As Gaussian Naive Bayes is comparatively easy to understand and quite simpler than the other two, this project work has been done using the same.

from sklearn.naive\_bayes import GaussianNB gnb = GaussianNB()

Naïve Bayes classifier depends on Bayes Theorem.

*Bayes theorem:*

P (Y/X1,.............. , Xn) = P(Y) P(X1, ......... ,Xn/Y)

.......................... P(X1, Xn)

Where,

Y is the class variable

X1 , X2, ......... , Xn are the dependent features.

**4.2. Data collection**

Data collection has been done from the internet to identify the disease here the real symptoms of the disease are collected i.e. no dummy values are entered. The dataset is collected from kagggle.com.[9]

This csv file contains 5000 rows and 133 columns, 132 columns for the unique symptoms. And last column for the disease class (40 unique disease classes).

Some rows of disease with their corresponding symptoms in the dataset –



**Performance analysis**

Confusion matrix-

A confusion matrix is basically a table that is used to describe the performance of a classification model on a group of test data that truth values are known. From the confusion matrix table, it is clearly seen that the Naïve Bayes algorithm is predicting all the diseases correctly within the test set.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| [[18, | 0, | | 0, | ..., | 0, | 0, | 0], |
| [ 0, | 30, | | 0, | ..., | 0, | 0, | 0], |
| [ 0, | 0, | | 24, | ..., | 0, | 0, | 0], |
| ..., | | | | | | | |
| [ 0, | | 0, | 0, | ..., | 6, | 0, | 0], |
| [ 0, | | 0, | 0, | ..., | 0, | 22, | 0], |
| [ 0, | | 0, | 0, | ..., | 0, | 0, | 34]] |

*Classification report-*

Classification report visualizes the precision, recall and f1 score of a model. 

*Accuracy score*: Our result showed a prediction accuracy of 1.0.

**Run Django project**

First make sure PostgreSQL and pgadmin4 is install in your system. then you have to manually create a DataBase instance on PostgreSQL named "predico", better use PgAdmin for that. make a new environment(recommended) and run...

- Run pip install dependencies

django==3.0.3

joblib==0.14.1

scikit-learn==0.21.3

psycopg2==2.8.4

- Run python manage.py makemigrations

- Run python manage.py migrate

- Run python manage.py runserver

- Navigate to http://127.0.0.1:8000/ in your browser

Db details

NAME': 'predico',

        'USER': 'postgres',

        'PASSWORD': '1997',

        'HOST': 'localhost'