A MINI PROJECT REPORT
ON

'Disease Prediction using Machine Learning'

Submitted by

Miss. Lakshmi Bhavsar (Roll No: 34)
Miss. Bhagyashri Niphade (Roll No: 07)
Miss. Laxmi Sonawane (Roll No: 19)

Under the guidance of

Prof. Vishal Patil

For the Subject

Laboratory Practice II (410247) -
Data Mining and Warehousing (410244 (D))

Submitted in partial fulfilment of the requirements
For the award of the degree of

Bachelor in Computer Engineering

MET Bhujbal Knowledge City

THE MET LEAGUE OF COLLEGES
AS SHARP AS YOU CAN GET

Institute of Engineering
Department of Computer Engineering

Academic Year 2021-22

i.

CONTENTS

1. Problem statement

Consider a labelled dataset belonging to an application domain. Apply suitable data pre-processing steps such as handling of null values, data reduction, and discretization. For prediction of class labels of given data instances, build classifier models using different techniques (minimum 3), analyse the confusion matrix and compare these models. Also apply cross validation while preparing the training and testing datasets.

For Example: Health Care Domain for predicting disease.

2. Abstract

This Machine Learning project is used to predict the disease based on the symptoms given by the user. It predicts using three different machine learning algorithms that is as follows:

- Random Forest Algorithm,
- Decision Tree,
- Naive Bayes Algorithm.

So, the output is accurate. It uses tkinter for GUI.

3. Introduction

This Machine Learning project is used to predict the disease based on the symptoms given by the user.

As we have also provided the dataset for the Testing the Symptoms of the patient or the user by the symptoms provided by them.

It predicts using three different machine learning algorithms that is as follows:

- **Random Forest Algorithm:**

Random Forest is a machine learning algorithm that belongs to the supervised learning technique. It can be used for Classification and Regression problems in Machine Learning.

As the name goes, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

How does Random Forest algorithm work?

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points.

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.
Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

- **Decision Tree:**

Just as Random Forest is a machine learning algorithm, Decision tree is also a machine learning algorithm falls under the category of supervised learning. They can also solve both regression and classification problems based on the situation.

Decision tree algorithm uses tree representation to solve the problem where each leaf node corresponds to a class label and attributes are represented on the node which is internal of the tree.

- **Naive Bayes Classifier Algorithm:**

Just as Random Forest Algorithm and Decision Tree, Naive Bayes algorithm is also a supervised learning algorithm, which is based on Bayes theorem.
It is mostly used in text classification that includes a high dimensional training dataset.
Naive Bayes is one of the easy and effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

It is used to predict on the basis of the probability of an object.
Some popular examples of Naive Bayes Algorithm are spam filtration,
Sentimental analysis, and classifying articles and etc.

Working of Naive Bayes' Classifier:

Working can be understood as mentioned below:

1. Convert the dataset provided b user into frequency tables.

2. Generate table by finding the probabilities of given features.

3. Now, use Bayes theorem to calculate the posterior probability.

4. Objective

- The Objective of this Mini Project is used to predict the disease on the symptoms given by the user or the patient.

- To Understand the Random Forest Algorithm.

- To Understand the Decision Tree Algorithm.

- To Understand the Naive Bayes' Algorithm.

5. Result

Following are the Screen shots of the Output:

1. The Output screen after running the Program:

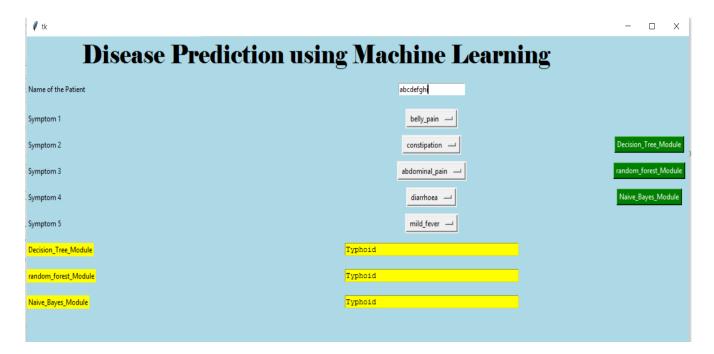2. After inserting the name of the Patient and the Symptoms of the Patient:

3. After Testing the First Algorithm that is the Decision Tree Algorithm.

4. After Testing the second Algorithm that is the Random Forest Algorithm.

5. After Testing the Third Algorithm that is the Naïve Bayes Algorithm. And the Final Output is as follows:

6. Conclusion

Hence, by using various Algorithm such as Random Forest Algorithm, Decision Tree, Naive Bayes Algorithm.
We have implemented the Mini Project on Disease Prediction using Machine Learning by taking the symptoms as input from User or the Patient.