MINI PROJECT REPORT

**Title –** Medical diagnosis of diseases based on adequate symptoms.

**Problem Statement –** Implement expert system to identify most probable disease patient is suffering from on basis of provided symptoms.

**Objective –**

1. To understand concept of Decision trees and Naïve Bayes classifier (MultinomialNB).
2. To compare performance of both algorithms by making them perform same task of prediction.

**Outcome -** After successfully completing this assignment, you should be able to

Display most probable disease patient is suffering on basis of passed symptoms.

**Software and Hardware requirements –**

1. Operating System : 64-bit Open source Linux or its derivative
2. Programming Languages: PYTHON/R

* **Algorithm :**

Step 1: Initialize value of K.

Step 2: Calculate distance between input sample and training samples.

Step 3: Sort the distances.

Step 4: Take top K- nearest neighbors.

Step 5: Apply simple majority.

Step 6: Predict class label for input sample.

* **Concepts related Theory :-**

K-Nearest Neighbor or KNN algorithm is part of supervised learning that has been used in many applications including data mining, statistical pattern recognition, and image processing. To identify the class of an input, the algorithm chooses the class to which the majority of the input’s k closest neighbors belong to. The KNN algorithm is considered as one of the simplest machine learning algorithms. KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. In pattern recognition, the k-Nearest Neighbors algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

• In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

• In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors. K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. A shortcoming of the k-NN algorithm is that it is sensitive to the local structure of the data. The algorithm has nothing to do with and is not to be confused with k-means, another popular machine learning technique.



**OPEN MPI**

The Open MPI Project is an open source [Message Passing Interface](http://www.mpi-forum.org/) implementation that is developed and maintained by a consortium of academic, research, and industry partners. Open MPI is therefore able to combine the expertise, technologies, and resources from all across the High Performance Computing community in order to build the best MPI library available. Open MPI offers advantages for system and software vendors, application developers and computer science researchers.

Features implemented or in short-term development for Open MPI include:

|  |  |
| --- | --- |
| * Full MPI-3.1 standards conformance * Thread safety and concurrency * Dynamic process spawning * Network and process fault tolerance * Support network heterogeneity * Single library supports all networks * Run-time instrumentation * Many job schedulers supported | * Many OS's supported (32 and 64 bit) * Production quality software * High performance on all platforms * Portable and maintainable * Tunable by installers and end-users * Component-based design, documented APIs * Active, responsive mailing list * Open source license based on the BSD license |

Open MPI is developed in a true open source fashion by a consortium of research, academic, and industry partners.

**Conclusion –** Hence, we have implemented an expert system to predict most probable disease patient is suffering from on basis of provided symptoms.

**Group Members:**

4411: Jayprakash Chawla

4412: Deepak Chaudhary