

**K. K.Wagh Institute of Engineering Education and Research, Nashik**

**Department of Computer Engineering**

**Model to categorize and predict symptoms in medicinal data**

Huge medical datasets available in various data repositories which are used for real world applications. To visualize the useful information stored in data warehouses, the Data Mining (DM) methods are enormously utilizedOn the way to categorize and predict symptoms in medicinal data, a variety of DM methods are utilized by different researchers. From many techniques of DM, classification is one of the main techniques. Decision trees can suffer from high variance which makes their results fragile to the specific training data used. Random Forest is an extension of bagging that in addition to building trees based on multiple samples of training data, it also constrains the features that can be used to build the trees, forcing trees to be different. This, in turn, can give a lift in performance. The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction.

Keywords: Data Mining, Machine Learning, Classification algorithms, Data mining, Decision trees, Partitioning algorithms, Training, Data models, Large Scale Database

Group ID: 8 Mini Project Guide

Class: BE – A Prof.K.P.Birla

Roll Nos: 29,30,31,32

Spandan Patel

Vedika Mujgule

Tanvi Joshi

Rashmi Purandare