



Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



**Department of Computer Science and Engineering (Data Science)**

**Subject: Machine Learning – I (DJ19DSC402)**

**AY: 2022-23**

**Jhanvi Parekh**

**60009210033**

**CSE(Data Science)**

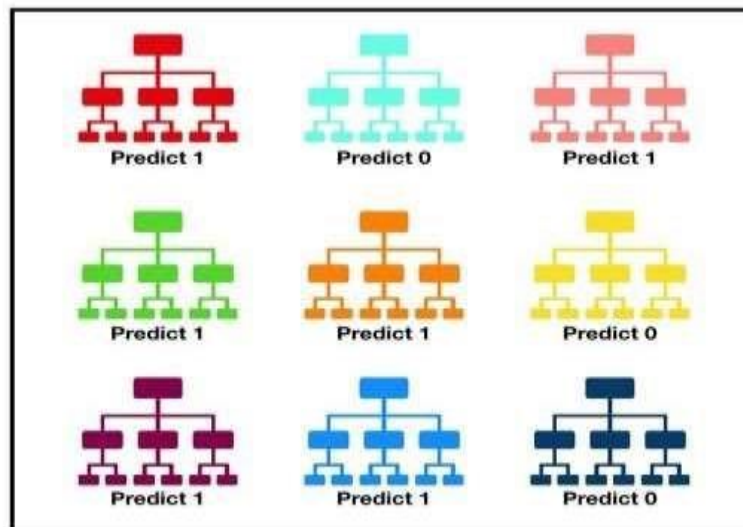
### **Experiment 6**

#### **(Random Forest)**

**Aim:** Implement Random Forest algorithm on given datasets and compare the results with Decision Tree classifiers for the same datasets.

#### **Theory:**

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction (see figure below).



Tally: Six 1s and Three 0s  
**Prediction: 1**

A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. The low correlation between models is the key. Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate

than any of the individual predictions. **The reason for this wonderful effect is that the trees protect each other from their individual errors** (as long as they don't constantly all err in the same direction). While

1



Shri Vile Parle Kelavani Mandal's  
**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**  
(Autonomous College Affiliated to the University of Mumbai)  
NAAC Accredited with "A" Grade (CGPA : 3.18)



### **Department of Computer Science and Engineering (Data Science)**

some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction. So the prerequisites for random forest to perform well are:

1. There needs to be some actual signal in our features so that models built using those features do better than random guessing.
2. The predictions (and therefore the errors) made by the individual trees need to have low correlations with each other.

#### **Lab Assignments to complete in this session:**

Use the given dataset and perform the following tasks:

##### **Dataset 1: IRIS.csv**

**Dataset 2: BehaviouralRiskFactorSurveillanceSystem.csv** (The objective of the BRFSS is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases in the adult population. Factors assessed by the BRFSS include tobacco use, health care coverage, HIV/AIDS knowledge or prevention, physical activity, and fruit and vegetable consumption. Data are collected from a random sample of adults (one per household) through a telephone survey. The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.)

1. Compare the results of decision tree and random forest classifier for dataset 1 and 2.
2. Compare the results of random forest with and without selecting important features only for building the classifier on dataset 2 and 3.

Colab File Link:

[https://colab.research.google.com/drive/15yZ0y4IDP\\_rS5dDMz7jPQEeWLuzvcWNp?usp=sharing](https://colab.research.google.com/drive/15yZ0y4IDP_rS5dDMz7jPQEeWLuzvcWNp?usp=sharing)

<https://colab.research.google.com/drive/1m2hUYoTVG6VrR5xRP8vgYSVoOU3dQy6k?usp=sharing>

[https://colab.research.google.com/drive/1LfztQs2YbraF7G0YcVwFBHf\\_caln-wAL?usp=sharing](https://colab.research.google.com/drive/1LfztQs2YbraF7G0YcVwFBHf_caln-wAL?usp=sharing)