

**COURSE CODE : CA5314**

**COURSE NAME : MACHINE LEARNING TECHNIQUES LABORATORY**

**PREDICTING THE QUALITY OF THE WINE USING  
MACHINE LEARNING ALGORITHMS**

Thenmozhi  
2021178054

## **Introduction:**

Red wine quality and style are highly influenced by the qualitative and quantitative composition of aromatic compounds having various chemical structures and properties and their interaction within different red wine matrices. The understanding of interactions between the wine matrix and volatile compounds and the impact on the overall flavor as well as on typical or specific aromas is getting more and more important for the creation of certain wine styles.

This experiment reports with a range of classic machine learning methods, including Decision Tree Classifier (DT), K-nearest neighbors' algorithm (KNN), Random Forest.

## **Decision Tree:**

A decision tree classifier is a systematic approach for multiclass classification. It poses a set of questions to the dataset (related to its attributes/features). The decision tree classification algorithm can be visualized on a binary tree. On the root and each of the internal nodes, a question is posed and the data on that node is further split into separate records that have different characteristics. The leaves of the tree refer to the classes in which the dataset is split. In the following code snippet, we train a decision tree classifier in scikit-learn.

## **KNN:**

KNN or k-nearest neighbors is the simplest classification algorithm. This classification algorithm does not depend on the structure of the data. Whenever a new example is encountered, its k nearest neighbors from the training data are examined. Distance between two examples can be the euclidean distance between their feature vectors. The majority class among the k nearest neighbors is taken to be the class for the encountered example.

## **Random Forest:**

Random forests are ensembles of decision trees: they consist of a bunch of independent decision trees, each of which is trained using only a subset of the features in our training set to ensure that they're learning to make their predictions in different ways. Their outputs are then pooled together using simple voting.

### **Evaluation Metrics**

1. Confusion Matrix
2. Accuracy

## **Dataset:**

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine has been taken to classify the quality of the wine as poor, good and excellent. The dataset has 12 attributes based on which the quality of the wine is being determined. Not all the attributes contribute significantly to the quality of the wine but most do.

**Link to the dataset:** <https://archive.ics.uci.edu/ml/datasets/wine+quality>

### **Class Type for Prediction**

If the quality lies between,

- 2 and 6 – quality is classified as “BAD”
- 6 and 8 – quality is classified as “GOOD”
- 8 and 9 – quality is classified as “EXCELLENT”

**Attribute Information**

<b>SNO</b>	<b>ATTRIBUTE</b>
1	type
2	fixed acidity
3	Volatile acidity
4	Citric acid
5	Residual sugar
6	chlorides
7	Free sulfur dioxide
8	Total sulfur dioxide
9	density
10	pH
11	sulphates
12	alcohol

## Results

Algorithm used	Accuracy
KNN	85.61%
Random forest	88.77%
Decision Tree	86.67%