**Problem:-**

Certain shop owners doesn’t know the demand of their customer due to which their profit slows down. This project is based on the text classification which classifies different types of wines according to the recommendation. It deals with wine classification which solves the problem of shop owners by helping them to already know about the demand of the wine and accordingly the stock will be updated. This will facilitate in increasing the profit of the shop owners.

**Objective:-**

The objective is to use modern and effective machine learning techniques like K Nearest Neighbors and Support Vector Machines which groups together the dataset and providing the comprehensive and generic approach for recommending wine to the customers on the basis of certain features.

**Learning algorithm used are as follows:-**

1. **K Nearest Neighbor**

K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection.

It is widely disposable in real-life scenarios since it is non-parametric, meaning; it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

The KNN Algorithm

1. Load the data

2. Initialize K to your chosen number of neighbors

3. For each example in the data

3.1 Calculate the distance between the query example and the current example from the data.

3.2 Add the distance and the index of the example to an ordered collection

4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances

5. Pick the first K entries from the sorted collection

6. Get the labels of the selected K entries

7. If regression, return the mean of the K labels

8. If classification, return the mode of the K labels

Advantages

The algorithm is simple and easy to implement.

There’s no need to build a model, tune several parameters, or make additional assumptions.

The algorithm is versatile. It can be used for classification, regression, and search (as we will see in the next section).

Disadvantages

The algorithm gets significantly slower as the number of examples and/or predictors/independent variables increase.

1. **Support Vector Machine**

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

 

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3.



Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.