



**S. B. JAIN INSTITUTE OF TECHNOLOGY,
MANAGEMENT & RESEARCH, NAGPUR.**

Practical No. 7

Aim: Implementation of K-Means algorithm for machine learning (On the same sample data as in program 8.) compare it with decision tree.

Name of Student: Shrutika Pradeep Bagdi

Roll No.: CS22130

Semester/Year: V/III

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

Department of Computer Science & Engineering, S.B.J.I.T.M.R., Nagpur

AIM: Implementation of K-Means algorithm for machine and compare it with decision tree.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- To be able to understand the concept of K-Means algorithm.
- To be able to implement K-Means algorithm for machine by comparing it with decision tree.

THEORY:

K-Means Algorithm: -

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabelled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if $K=2$, there will be two clusters, and for $K=3$, there will be three clusters, and so on. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The algorithm takes the unlabelled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

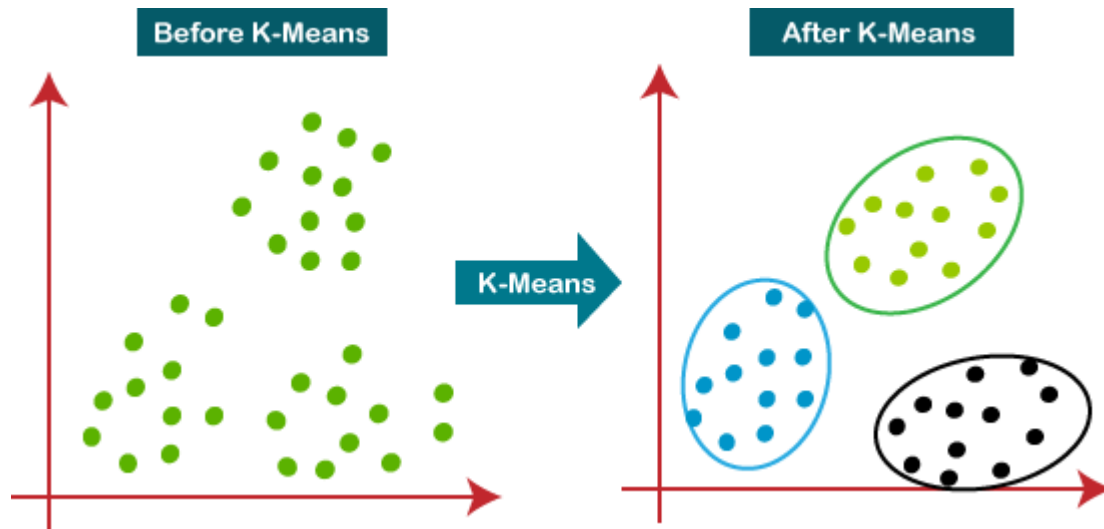
The k-means clustering algorithm mainly performs two tasks:

Determines the best value for K- center points or centroids by an iterative process.

Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Hence each cluster has datapoints with some commonalities, and it is away from other clusters.

The below diagram explains the working of the K-means Clustering Algorithm:



How does the K-Means Algorithm Work?

The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be other from the input dataset).

Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready.

Decision Tree is a supervised learning algorithm; k-means clustering is an unsupervised learning algorithm.

ScreenShot:

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer: Choose **EM** -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Cluster mode

- ☐ Use training set
- ☐ Supplied test set (Set...)
- ☐ Percentage split % 66
- ☒ Classes to clusters evaluation (Nom) class
- ☒ Store clusters for visualization

Ignore attributes: []

Start Stop

Result list (right-click for options)

- 03:23:33 - SimpleKMeans
- 03:23:39 - EM**

Clusterer output

=== Run information ===

Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation: iris
Instances: 150
Attributes: 5
sepalength
sepalwidth
petallength
petalwidth

Ignored: class

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

EM

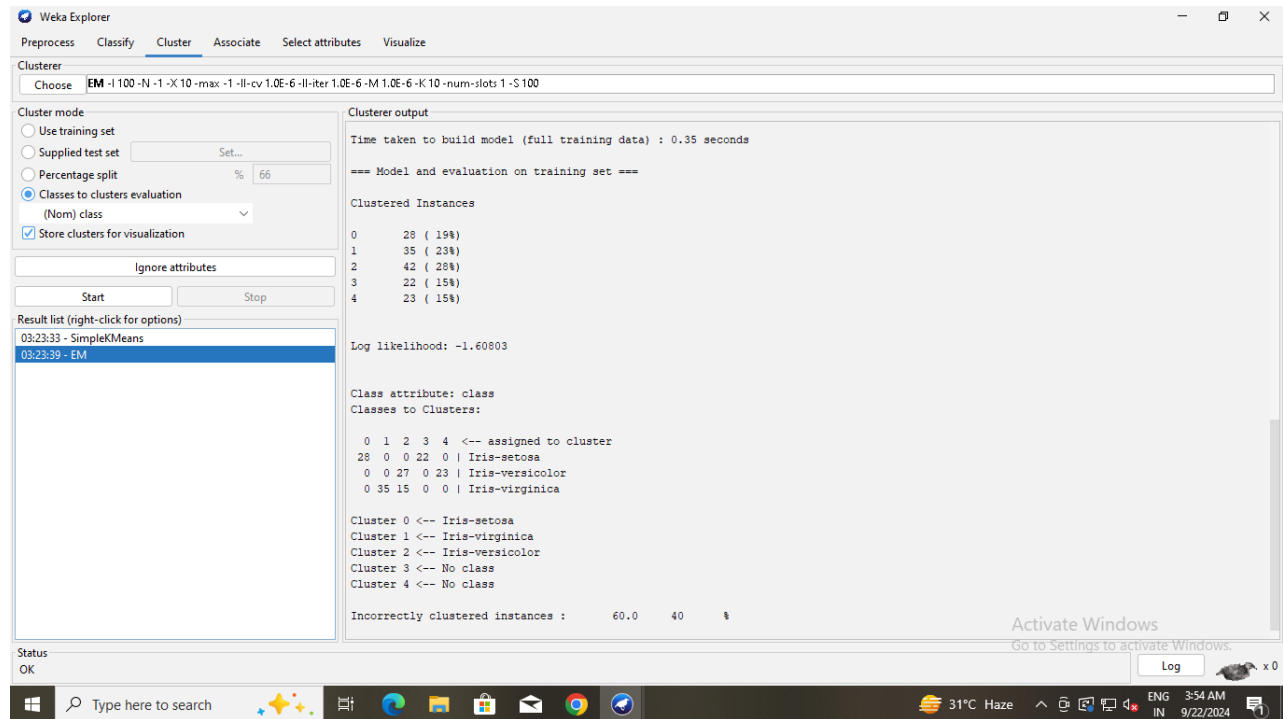
Number of clusters selected by cross validation: 5
Number of iterations performed: 16

Attribute	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
sepalength	mean 4.7748	6.0585	6.1613	5.2823	5.5432
	std. dev. 0.2405	0.5228	0.4138	0.2407	0.3159

Status: OK

Activate Windows: Go to Settings to activate Windows.





CONCLUSION:

Thus , successfully implement the K-Means algorithm for machine learning .

DISCUSSION QUESTIONS:

- Q.1) What is K-means algorithm and how it works?
- Q.2) Why K-means is a clustering algorithm?
- Q.3) What are the properties of k-means algorithm?
- Q.4) Why K-means algorithm is used?

REFERENCES:

- <https://byjus.com/maths/magic-square/>
- <https://www.geeksforgeeks.org/magic-square/>