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

Practical No. 8

Aim: Implementation of K-Means Clustering and Data Mining Techniques using Orange3-3.31.1 Toolkit.

Name of Student: Shrutika Pradeep Bagdi

Roll No. : CS22130

Semester/Year : 6th sem/3rd year

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

AIM: Implementation of K-Means Clustering and Data Mining Techniques using Orange3-3.31.1 Toolkit.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

- By using new data mining tool Orange, students will explore its functionalities and capabilities in data preprocessing, feature selection, modeling, evaluation, and visualization.
- Through hands-on exercises, students will apply a variety of data mining algorithms available in Orange, including decision trees, support vector machines, k-nearest neighbors, k-means clustering, and association rule mining, among others.
- Through project-based learning activities, students will develop critical thinking skills by formulating hypotheses, designing experiments, analyzing results, and making data-driven decisions to solve complex analytical problems.

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement: Computer System with high configurations

Software Requirement: Orange 3.38.1 for Windows

THEORY:

ORANGE DATA MINING

Orange is a C++ core object and routines library that incorporates a huge variety of standard and non-standard machine learning and data mining algorithms. It is an open-source data visualization, data mining, and machine learning tool. Orange is a scriptable environment for quick prototyping of the latest algorithms and testing patterns. It is a group of python-based modules that exist in the core library. It implements some functionalities for which execution time is not essential, and that is done in Python.

It incorporates a variety of tasks such as pretty-print of decision trees, bagging and boosting, attribute subset, and many more. Orange is a set of graphical widgets that utilizes strategies from the core library and orange modules and gives a decent user interface. The widget supports digital-based communication and can be gathered together into an application by a visual programming tool called an orange canvas.

All these together make an orange an exclusive component-based algorithm for data mining and machine learning. Orange is proposed for both experienced users and analysts in data mining and machine learning

who want to create and test their own algorithms while reusing as much of the code as possible, and for those simply entering the field who can either write short python contents for data analysis.

The objective of Orange is to provide a platform for experiment-based selection, predictive modeling, and recommendation system. It primarily used in bioinformatics, genomic research, biomedicine, and teaching. In education, it is used for providing better teaching methods for data mining and machine learning to students of biology, biomedicine, and informatics.

Orange core objects and Python modules incorporate numerous data mining tasks that are far from data preprocessing for evaluation and modeling. The operating principle of Orange is cover techniques and perspective in data mining and machine learning. For example, Orange's top-down induction of decision tree is a technique build of numerous components of which anyone can be prototyped in python and used in place of the original one. Orange widgets are not simply graphical objects that give a graphical interface for a specific strategy in Orange, but it includes an adaptable signaling mechanism that is for communication and exchange of objects like data sets, classification models, learners, objects that store the results of the assessment. All these ideas are significant and together recognize Orange from other data mining structures.

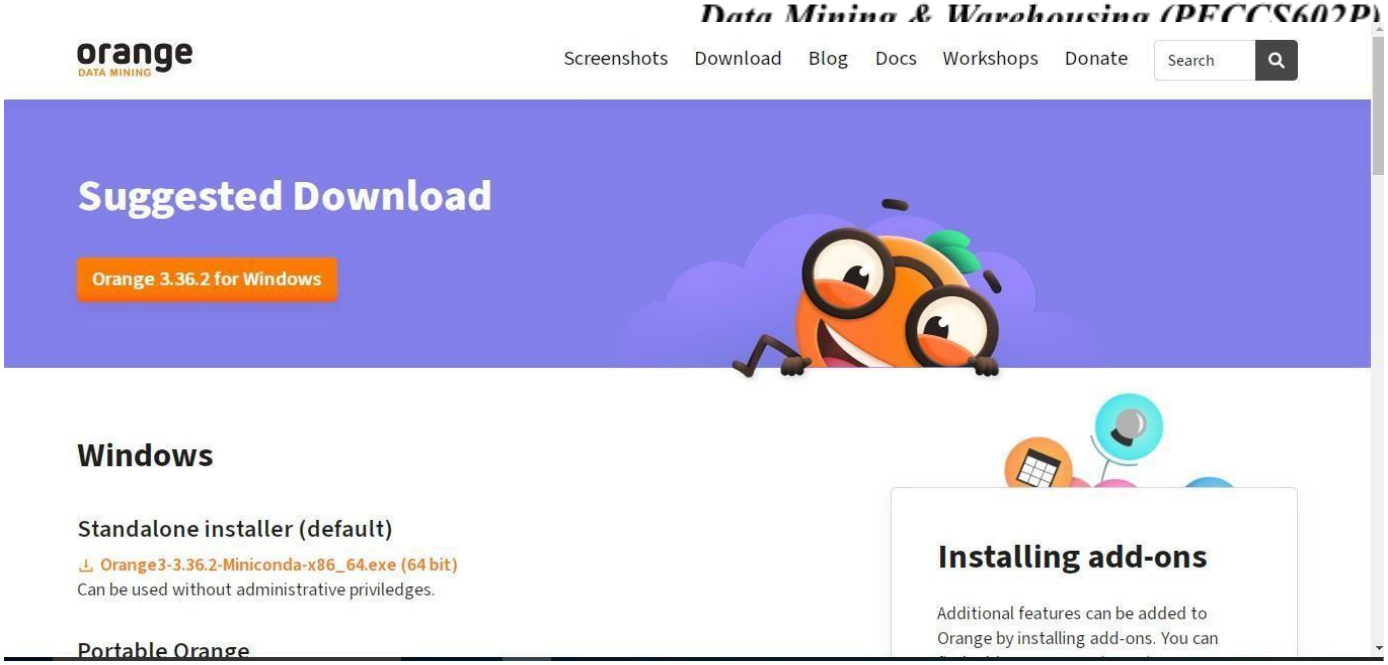
ORANGE WIDGETS:

Orange widgets give us a graphical user interface to orange's data mining and machine learning techniques. They incorporate widgets for data entry and preprocessing, classification, regression, association rules and clustering a set of widgets for model assessment and visualization of assessment results, and widgets for exporting the models into PMML.

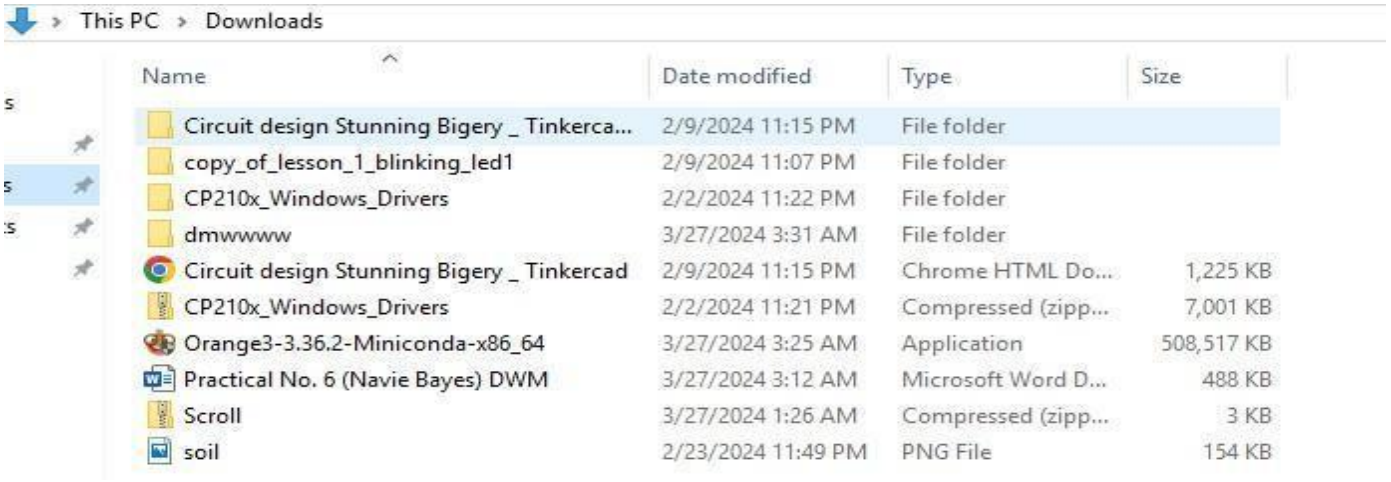
Widgets convey the data by tokens that are passed from the sender to the receiver widget. For example, a file widget outputs the data objects, that can be received by a widget classification tree learner widget. The classification tree builds a classification model that sends the data to the widget that graphically shows the tree. An evaluation widget may get a data set from the file widget and objects.

Download and install Orange 3.38.1 :

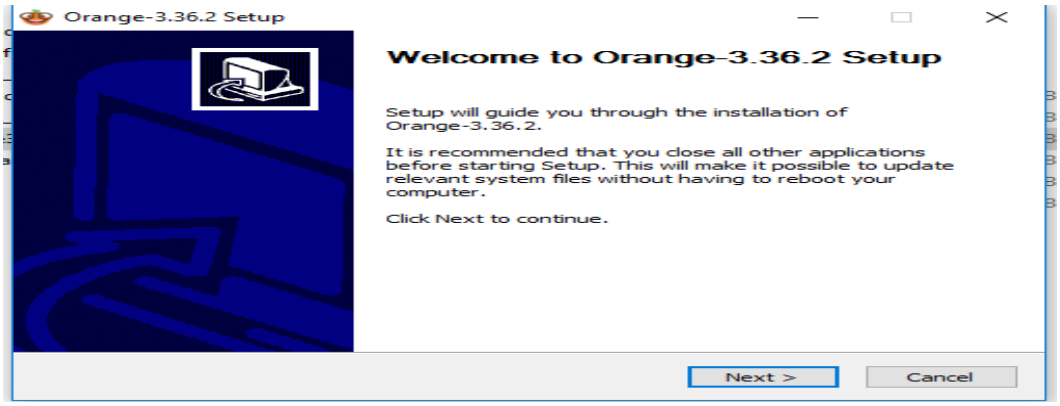
Step-1 : Install Orange 3.38.1 for windows from : https://orangedatamining.com/download/



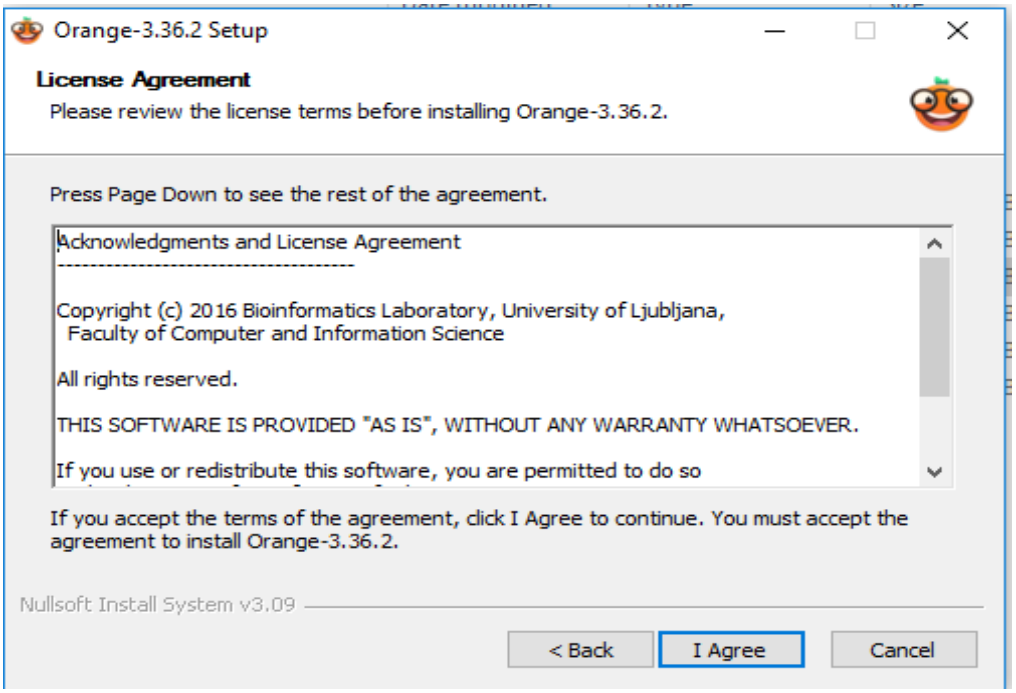
Step-2 : After installation open the file



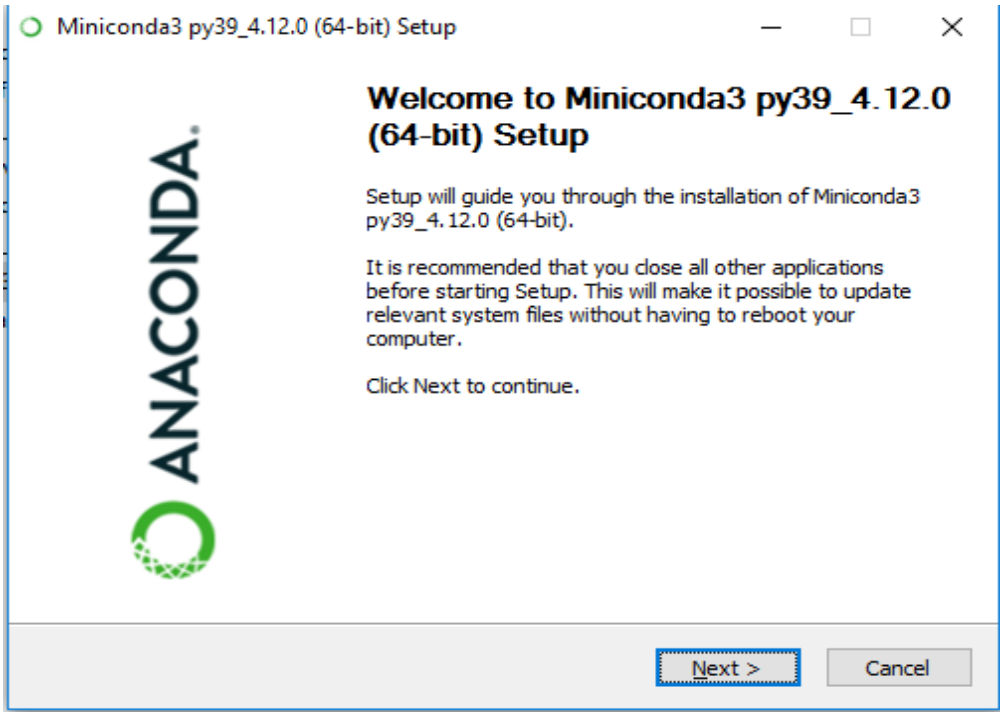
Step-3 : Click on "Next"



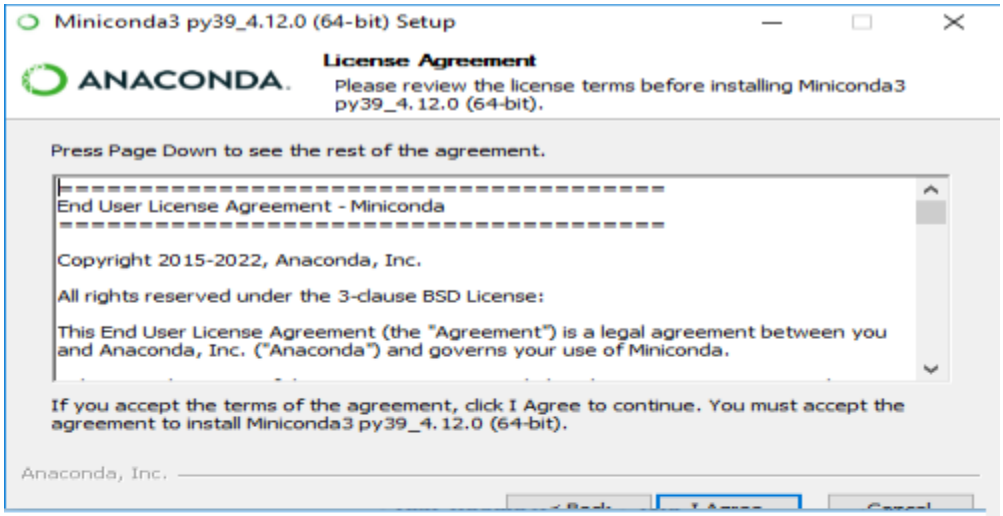
Step-4 : Click "I Agree"

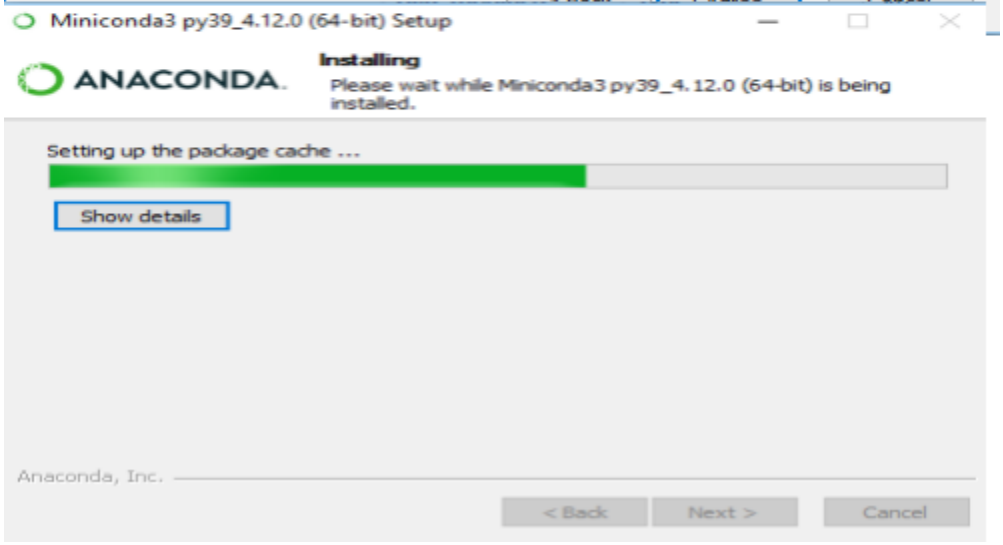


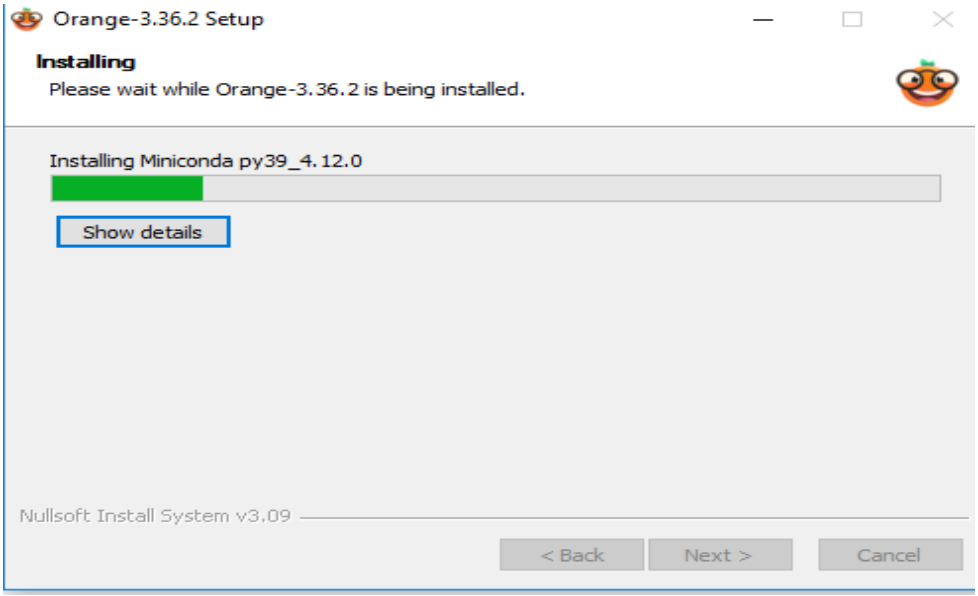
Step-5 : Click "Next"



Step-6 : Click "I Agree"


Miniconda3 py39_4.12.0 (64-bit) Setup
ANACONDA.
License Agreement
Please review the license terms before installing Miniconda3 py39_4.12.0 (64-bit).
Press Page Down to see the rest of the agreement.
End User License Agreement - Miniconda
Copyright 2015-2022, Anaconda, Inc.
All rights reserved under the 3-clause BSD License:
This End User License Agreement (the "Agreement") is a legal agreement between you and Anaconda, Inc. ("Anaconda") and governs your use of Miniconda.
If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install Miniconda3 py39_4.12.0 (64-bit).
Anaconda, Inc.


Miniconda3 py39_4.12.0 (64-bit) Setup
ANACONDA.
Installing
Please wait while Miniconda3 py39_4.12.0 (64-bit) is being installed.
Setting up the package cache ...
Show details
Anaconda, Inc.
< Back
Next >
Cancel


Orange-3.36.2 Setup
Installing
Please wait while Orange-3.36.2 is being installed.
Installing Miniconda py39_4.12.0
Show details
Nullsoft Install System v3.09
< Back
Next >
Cancel

PROCEDURE:

Step 1: Install Orange Mining Tool

- If you haven't already installed Orange, download and install it from the official website: https://orange.biolab.si/

Step 2: Launch Orange Mining Tool

- Open Orange Mining Tool from your applications or start menu.

Step 3: Load Paintdata

- Click on 'File' in the menu bar.
- Select 'Open' and navigate to the location where your Paintdata dataset is saved.
- Choose the Paintdata file and click 'Open' to load it into Orange.

Step 4: Preprocess Data (if needed)

- If your dataset requires preprocessing (e.g., handling missing values, normalization), use Orange's built-in functionalities to preprocess the data accordingly.

Step 5: Perform k-nearest neighbors (kNN)

- Drag the 'kNN' widget from the 'Classify' section onto the canvas.
- Connect the output of the Paintdata dataset to the 'kNN' widget.
- Set the desired parameters such as the number of neighbors (k) in the widget's settings.
- Optionally, you can split your data into training and testing sets using the 'Test & Score' widget to evaluate the model's performance.

Step 6: Visualize kNN Results

- Drag the 'Scatter Plot' widget from the 'Visualize' section onto the canvas.
- Connect the output of the 'kNN' widget to the 'Scatter Plot' widget.
- Choose the features you want to visualize by selecting them in the 'Scatter Plot' widget's settings.
- The scatter plot will display the distribution of data points and how the kNN algorithm classifies them.

Step 7: Perform k-means Clustering

- Drag the 'k-Means' widget from the 'Clustering' section onto the canvas.
- Connect the output of the Paintdata dataset to the 'k-Means' widget.
- Set the desired parameters such as the number of clusters (k) in the widget's settings.

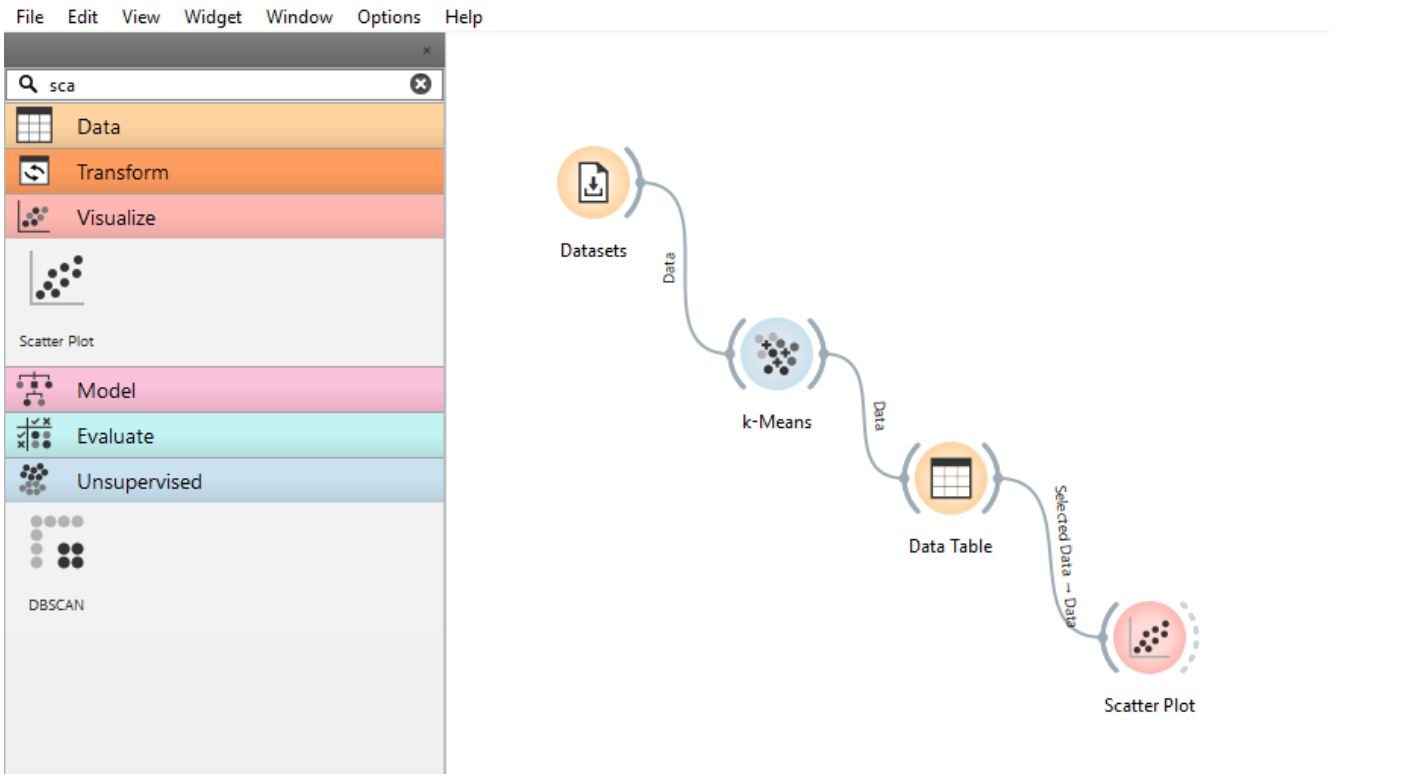
Step 8: Visualize k-means Clustering Results

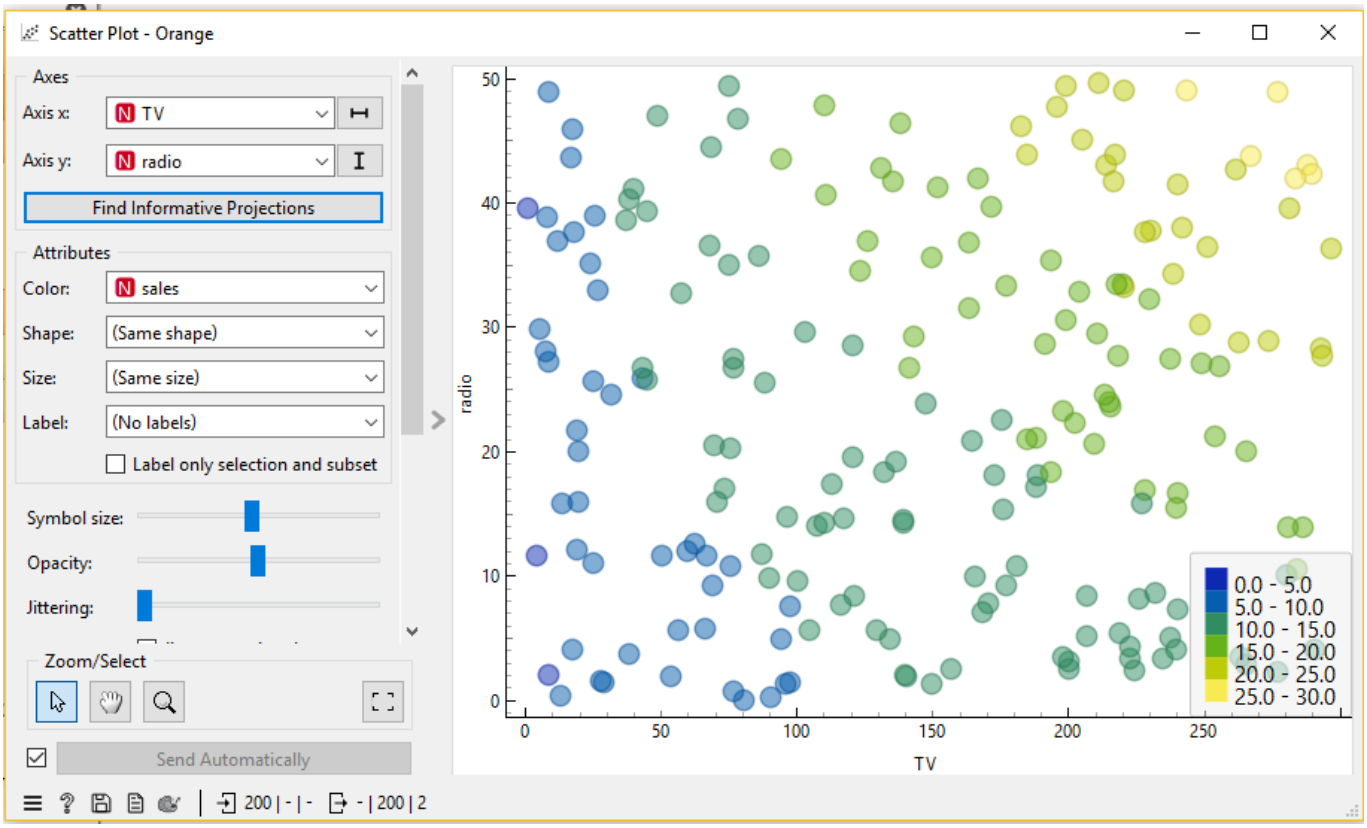
- Drag another 'Scatter Plot' widget from the 'Visualize' section onto the canvas.
- Connect the output of the 'k-Means' widget to the new 'Scatter Plot' widget.
- Choose the features you want to visualize by selecting them in the 'Scatter Plot' widget's settings.
- The scatter plot will display the clusters formed by the k-means algorithm.

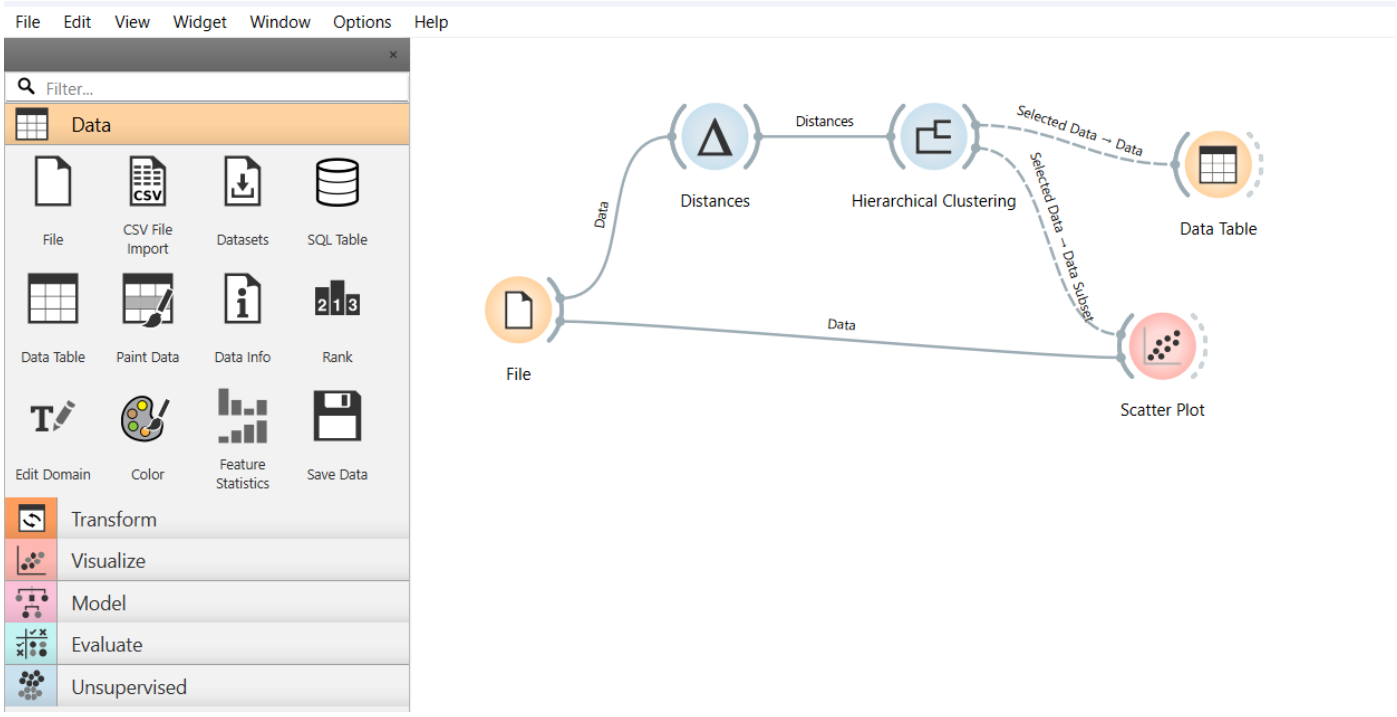
Step 9: Analyze and Interpret Results

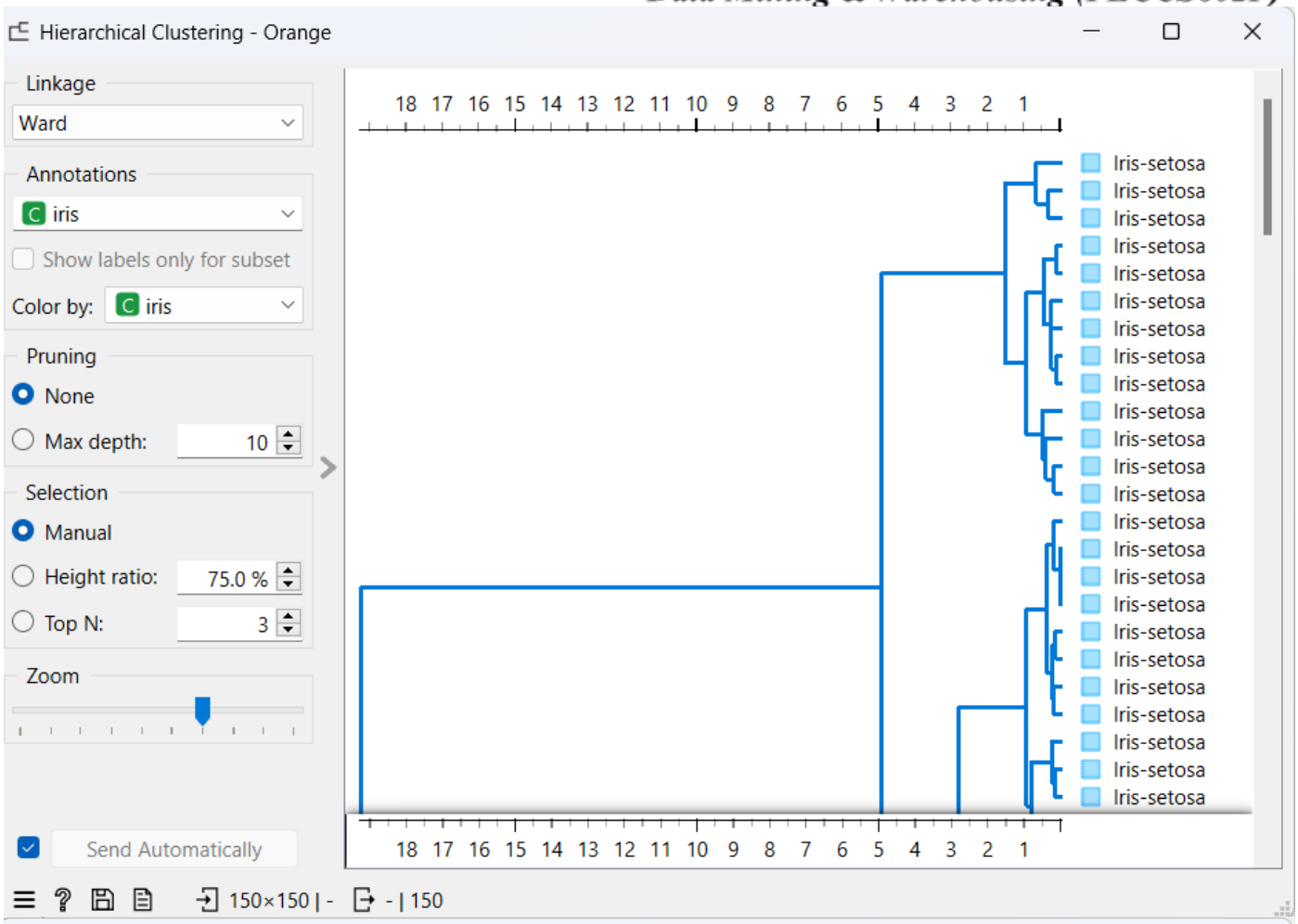
- Analyze the results of both kNN classification and k-means clustering.
- Interpret the visualizations to understand how the algorithms have classified or clustered the data.
- Make any necessary adjustments to parameters or preprocessing steps to improve the results if needed.

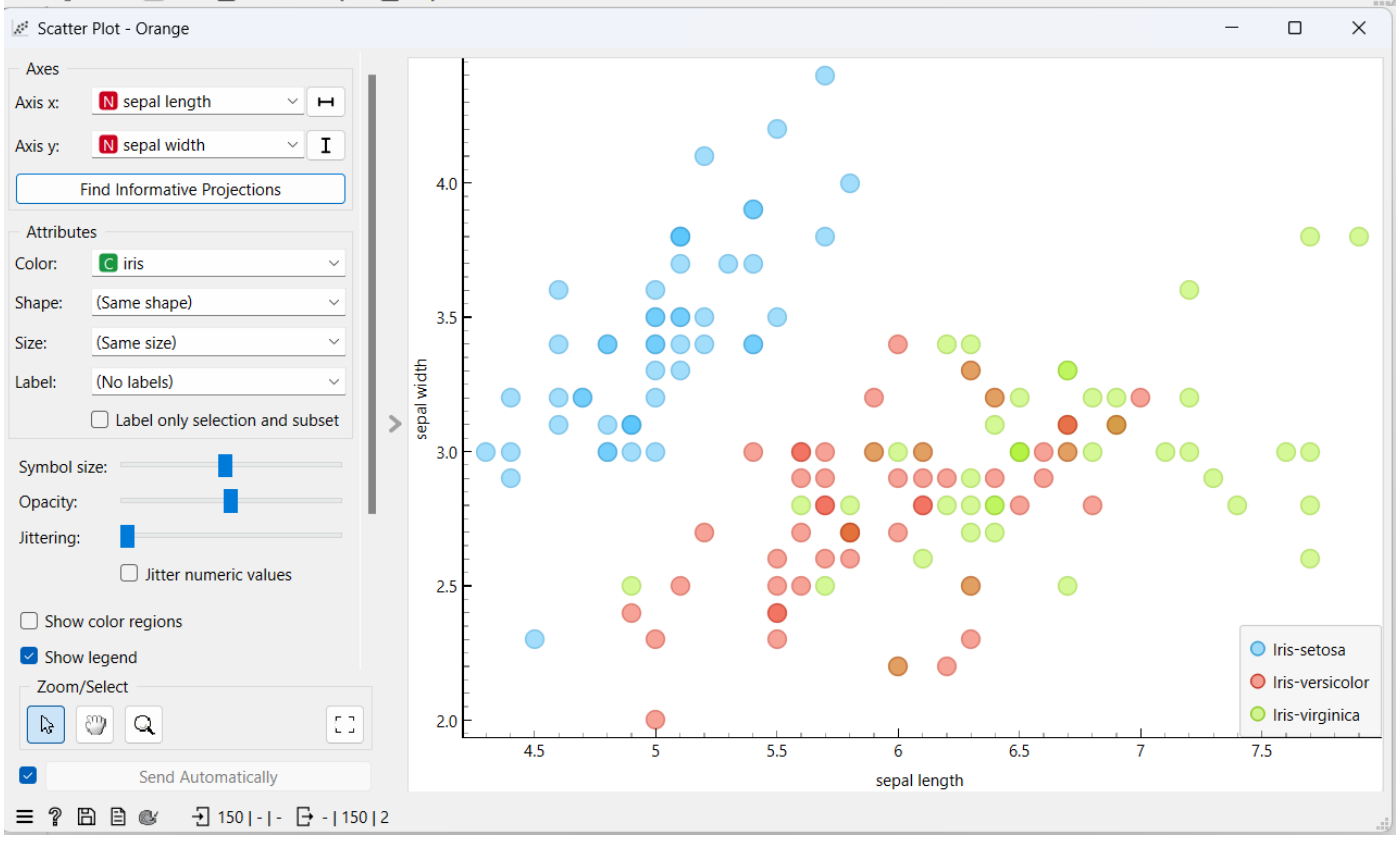
OUTPUT (SCREENSHOTS) :




Scatter Plot - Orange
Axes
Axis x: TV
Axis y: radio
Find Informative Projections
Attributes
Color: sales
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Zoom/Select
Send Automatically
200 | - | - - | 200 | 2
radio
TV
0.0 - 5.0
5.0 - 10.0
10.0 - 15.0
15.0 - 20.0
20.0 - 25.0
25.0 - 30.0


File Edit View Widget Window Options Help
Filter...
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Visualize
Model
Evaluate
Unsupervised
Distances
Hierarchical Clustering
Data Table
File
Scatter Plot
Data
Distances
Selected Data → Data
Selected Data → Data Subset
Data


Hierarchical Clustering - Orange
Linkage
Ward
Annotations
iris
Show labels only for subset
Color by: iris
Pruning
None
Max depth: 10
Selection
Manual
Height ratio: 75.0 %
Top N: 3
Zoom
Send Automatically
18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
Iris-setosa
150×150 | -
- | 150


Scatter Plot - Orange
Axes
Axis x: sepal length
Axis y: sepal width
Find Informative Projections
Attributes
Color: iris
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Jitter numeric values
Show color regions
Show legend
Zoom/Select
Send Automatically
sepal width
sepal length
4.0
3.5
3.0
2.5
2.0
4.5
5
5.5
6
6.5
7
7.5
Iris-setosa
Iris-versicolor
Iris-virginica
150 | - | -
- | 150 | 2

CONCLUSION:

DISCUSSION AND VIVA VOCE:

1. What are the different tools for data mining?
2. Can kNN and k-means clustering help identify distinct groups within the Paintdata dataset?
3. What type of data does Orange support.
4. Explain Orange Widgets.

REFERENCE:

https://orangedatamining.com/download/

https://orangedatamining.com/

https://orangedatamining.com/getting-started/

https://www.javatpoint.com/orange-data-mining