**DATA WAREHOUSING AND DATA MINING LAB**

**Experiment 7: Build a Decision Tree using ID3 or C4.5 algorithm and visualize the result**

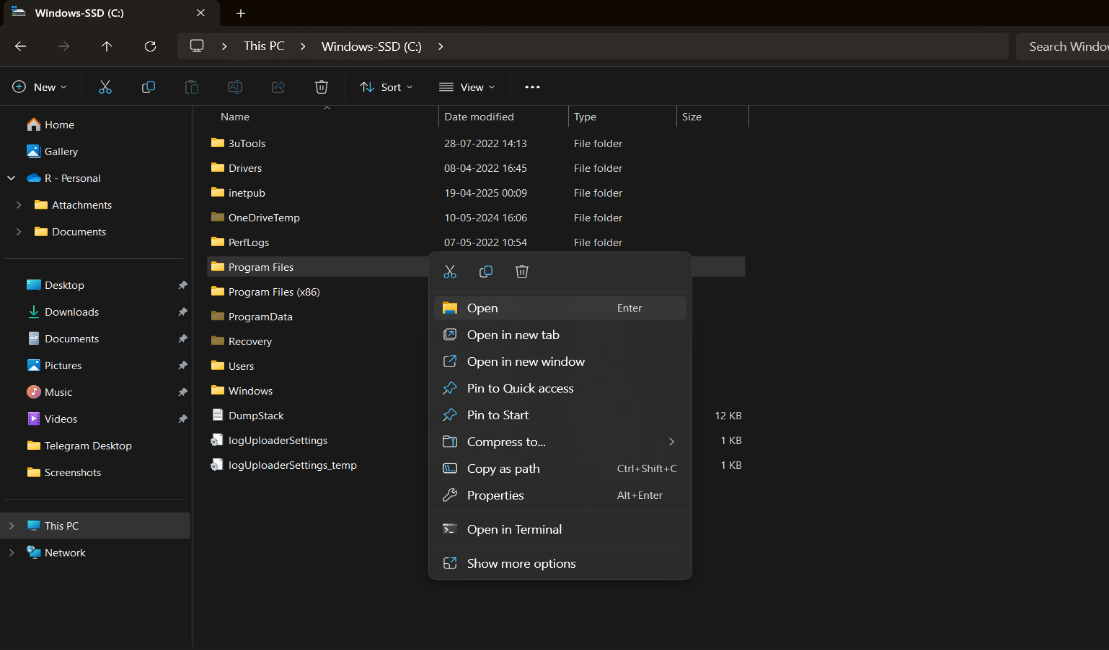
**Aim:**

To construct and visualize a decision tree using the C4.5 algorithm (J48 in WEKA) and analyze its decision-making process for classification.

**Description:**

Decision trees are flowchart-like models where internal nodes represent tests on attributes, branches represent outcomes, and leaf nodes represent classes. C4.5 uses **information gain ratio** to choose the best attribute for splitting. It is widely used because it is easy to interpret and can handle both numerical and categorical data.

**Steps to Run in WEKA :  
Step 1:**  
Open **File Explorer**  
Select **This PC** 🡪 Open the **C drive** 🡪 Open **Program Files**

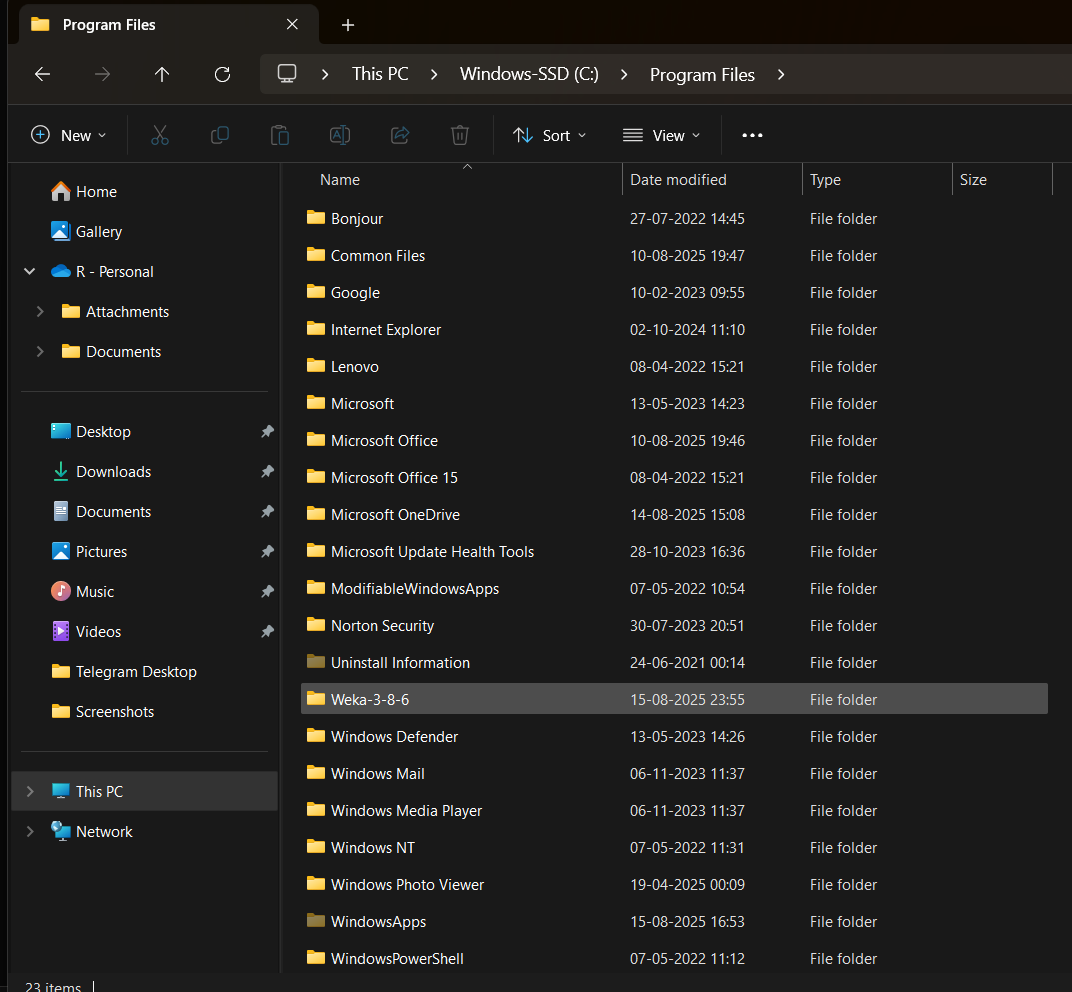


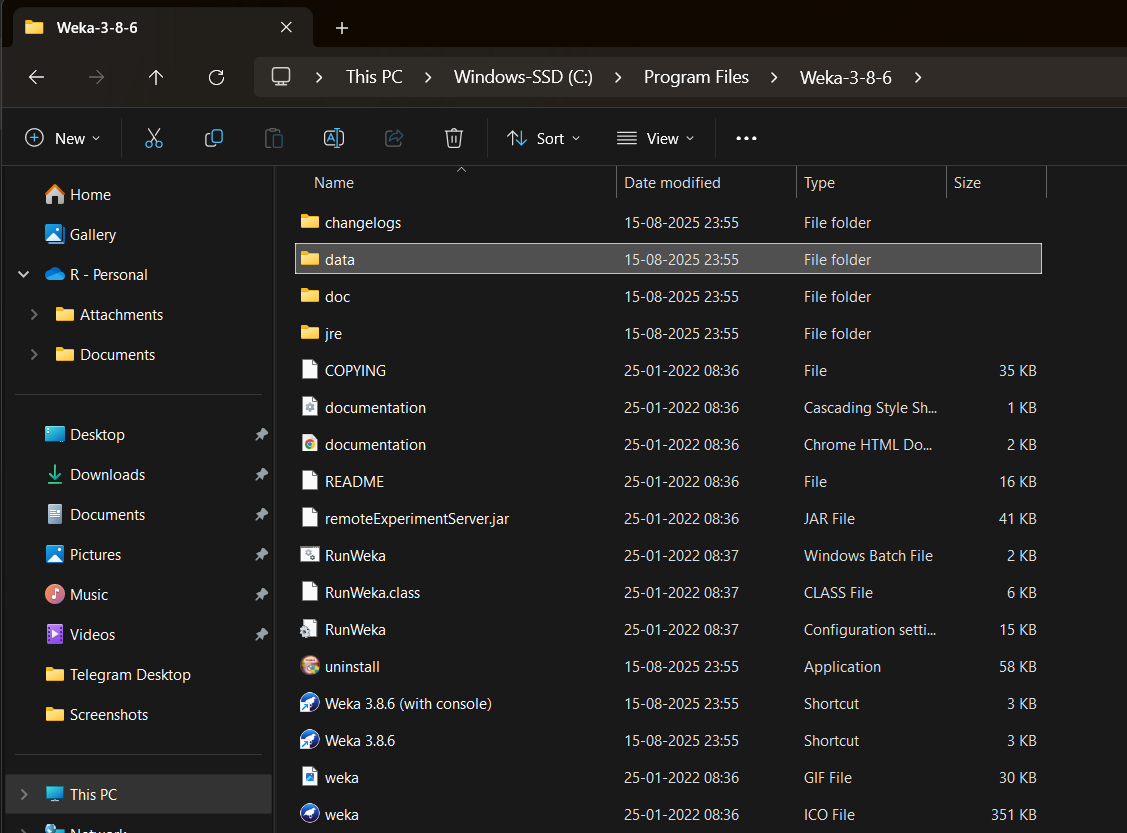
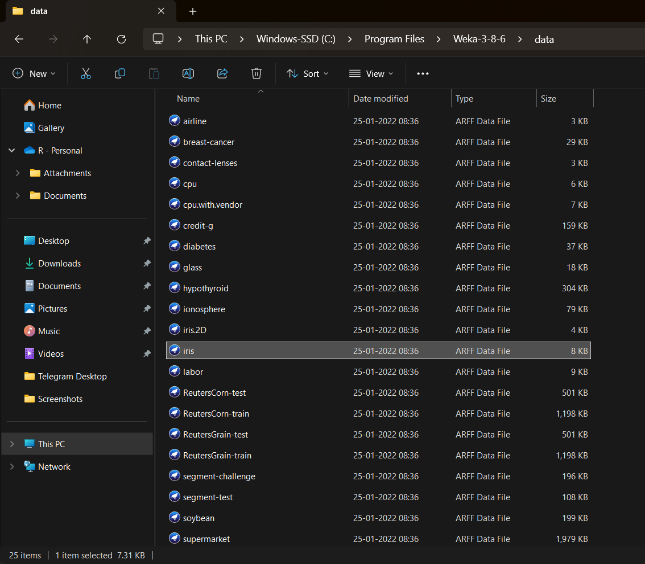
**Step 2 :**

Open the **WEKA 3.8.6** folder 🡪 Open the **Data** folder 🡪 Select the **Iris** dataset 🡪 Double-click on it

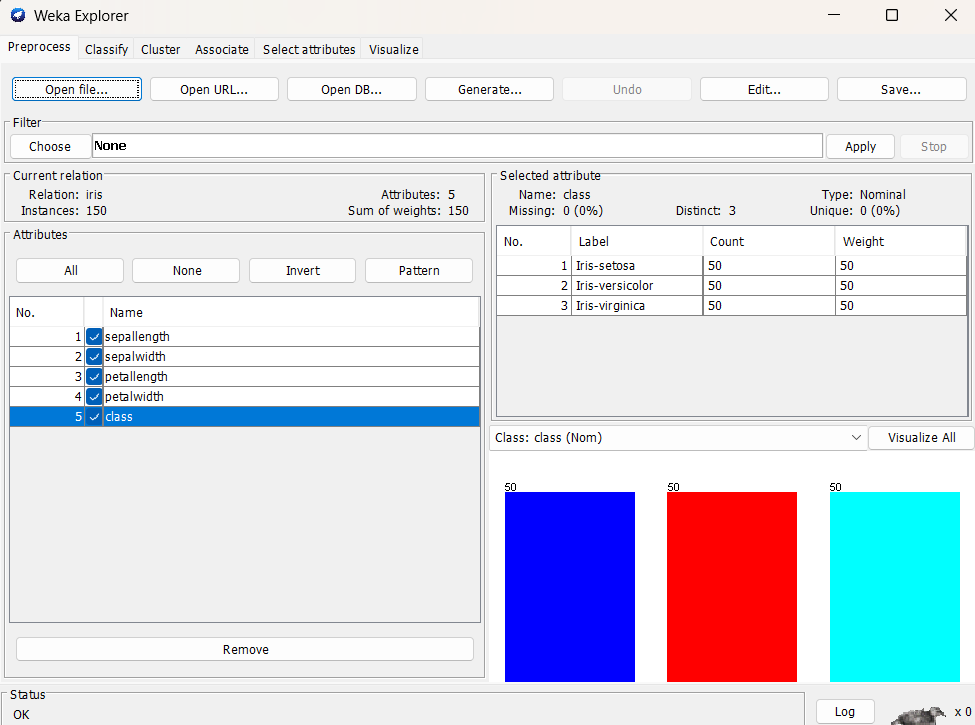
### **Iris Dataset – Information**

* **Introduced by**: Ronald A. Fisher in 1936 (in his paper on discriminant analysis).
* **Purpose**: Used for pattern recognition and classification tasks; one of the most famous benchmark datasets in machine learning.
* **Instances**: 150 flower samples.
* **Attributes**: 4 numerical features (in centimeters):
  1. Sepal length
  2. Sepal width
  3. Petal length
  4. Petal width
* **Class Labels (Species)**: 3 classes, each with **50 samples**:
  1. Iris-setosa
  2. Iris-versicolor
  3. Iris-virginica



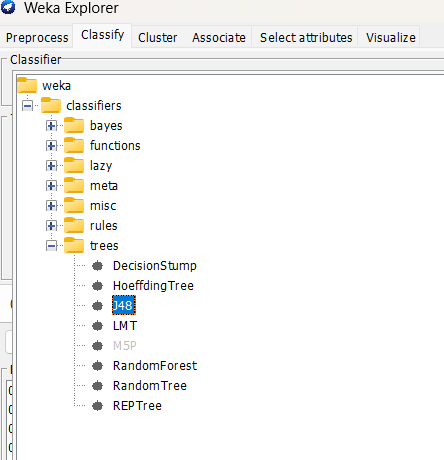
 

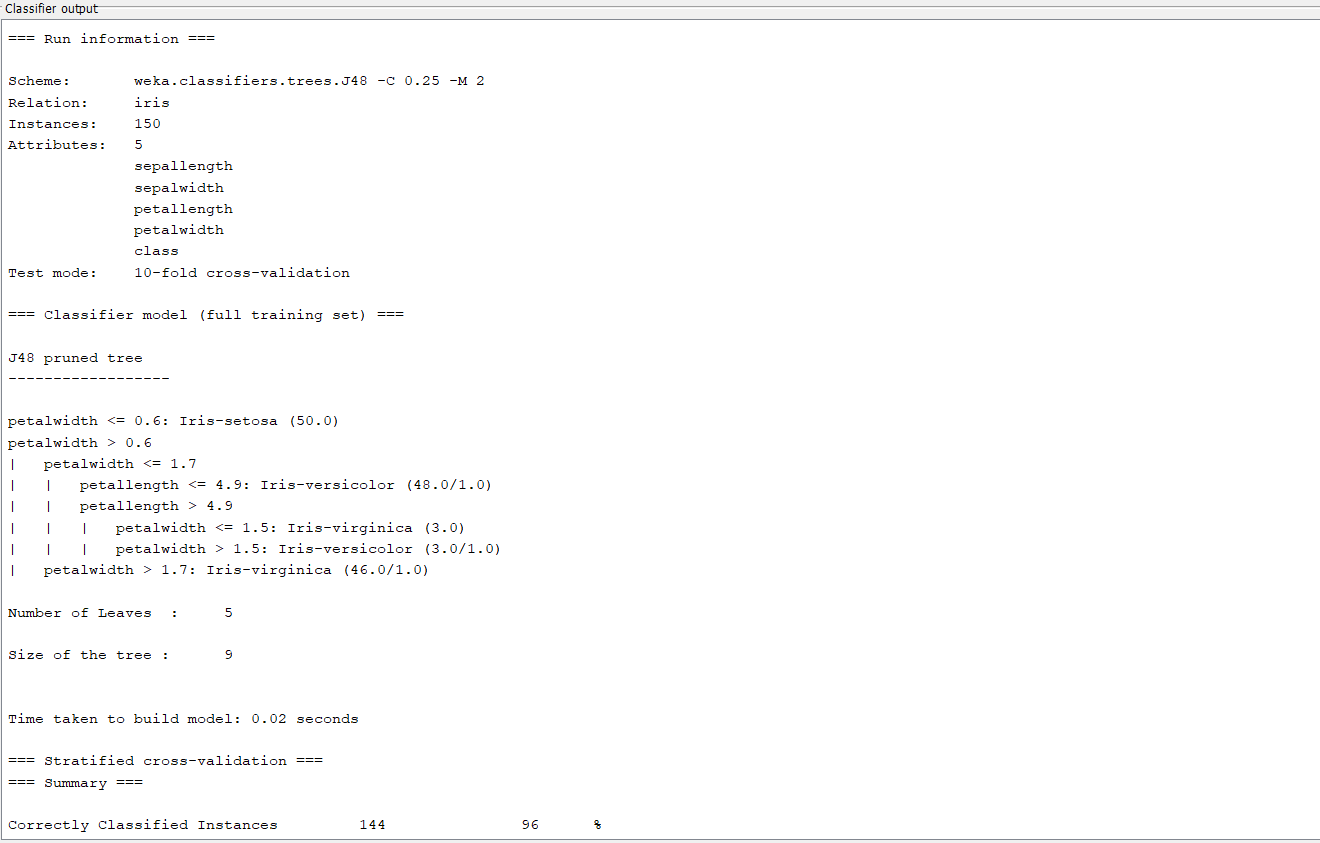
**Step 3:**  
 After WEKA opens, select all the parameters displayed there, such as **Sepallength, Sepalwidth, Petallength, Petalwidth,** and **Class**



**Step 4:**

Select **Classify** 🡪 Click the **Choose** option 🡪 In the **Classifiers** section, select **Lazy** 🡪  
Then, select **J48** from the options

 **Step 5:** Click on **Start**, and the output will be as follows:



### **J48 Decision Tree – Summary**

* **Dataset**: Iris (150 instances, 5 attributes) ,( The data used to train and test the model)
* **Evaluation**: 10-fold cross-validation (The method used to test the model’s performance)
* **Accuracy**: **96% (144/150 correct) (**The percentage of correctly classified instances)
* **Tree size**: 9 nodes, 5 leaves (The total number of nodes (decision points + leaves) in the decision tree)
* **Build time**: 0.02 sec (Time taken by the algorithm to construct the model)

**Decision Tree Rules (Simplified):**

* If **petalwidth ≤ 0.6** → Iris-setosa
* Else if **petalwidth > 1.7** → Iris-virginica
* Else → Mostly Iris-versicolor (with some splits using petallength)

**Result:**

* All setosa perfectly classified.
* Few misclassifications happen between versicolor and virginica.
* Performance is **same as Naive Bayes (96%)**, but here you also get an **interpretable tree model**.

**Conclusion** : J48 achieved 96% accuracy; simple tree rules clearly separate setosa and mostly distinguish versicolor from virginica.