

## Week 4 Practical Lab: Decision tree classifier using KNIME

### Useful resources:

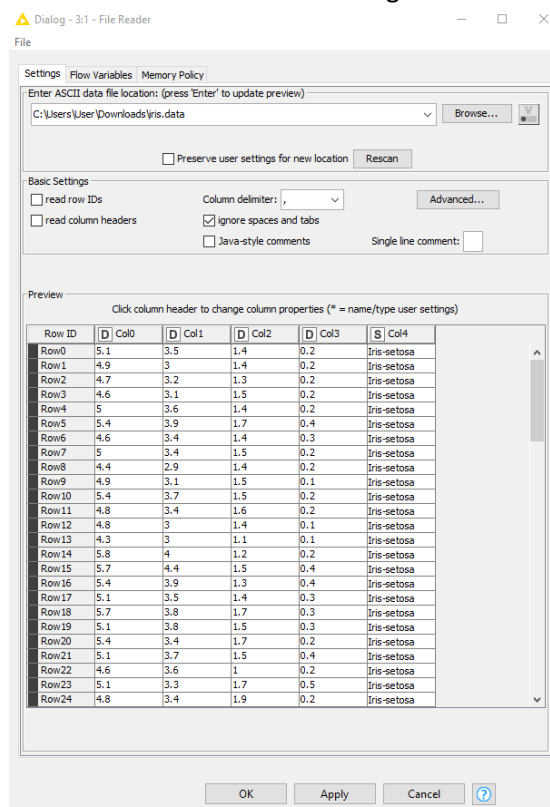
- **KNIME Analytics Platform** download link:  
<https://www.knime.com/downloads/download-knime>  
(It is recommended to download the installer version 561 MB for Windows)
- **Example Data set, Iris Data** <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/>

### Objective:

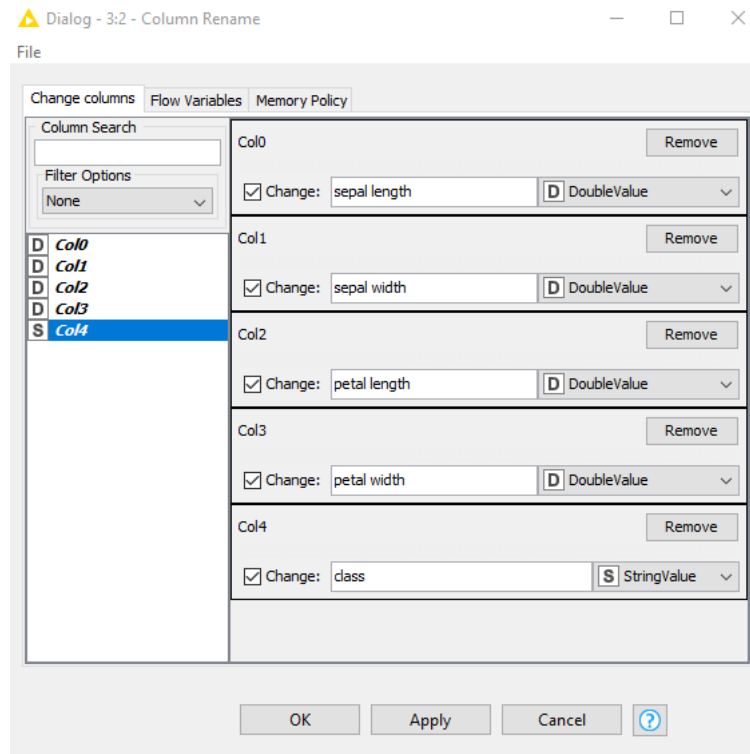
- Implement decision tree classifier using KNIME on a sample data set

### Instructions:

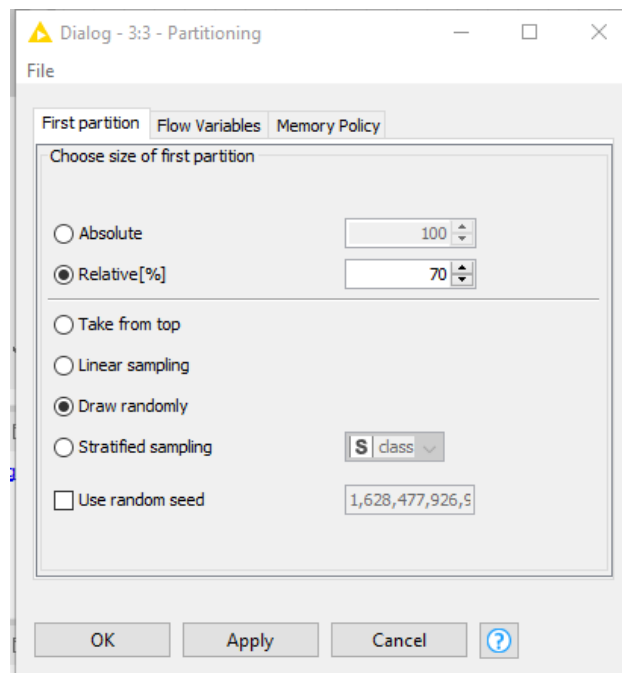
1. Open the KNIME software installed on your computer and create a new workflow
2. Select a data set of your choice from the given link  
(<https://archive.ics.uci.edu/ml/index.php>) and open the file using the appropriate node reader.
3. Execute the **File Reader** node and observe data through the **File Table** option.



4. To rename the column names, use the “Column Rename” node and rename as follows:



5. Now, to split the data set into training and testing, use the “Partitioning” node and configure as follows:



6. Next, connect the “**Decision Tree Learner**” node with the first output port of the **Partitioning** node and configure the learner node. Then execute the node and observe the decision tree view.

Dialog - 3:4 - Decision Tree Learner

File

Options PMMLSettings Flow Variables

General

Class column

Quality measure

Pruning method

☒ Reduced Error Pruning

Min number records per node

Number records to store for view

☒ Average split point

Number threads

☒ Skip nominal columns without domain information

Root split

☐ Force root split column

Root split column

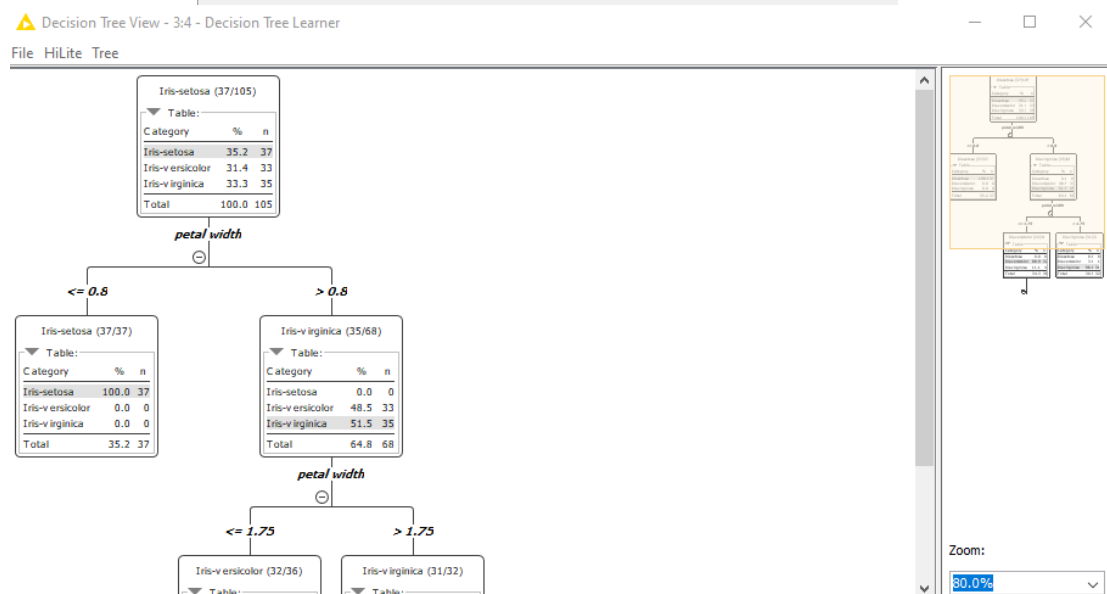
Binary nominal splits

☐ Binary nominal splits

Max #nominal

☐ Filter invalid attribute values in child nodes

OK Apply Cancel ?



- To test the model, take the “Decision tree predictor” node and connect it with the second output port of the Partitioning node and decision tree learner node. Once the connections are made, configure the predictor node.

**Dialog - 3:5 - Decision Tree Predictor**

File

Options Flow Variables Memory Policy

Maximum number of stored patterns for HiLite-ing: 10,000

☒ Change prediction column name

Prediction (class)

☐ Append columns with normalized class distribution

Suffix for probability columns

OK Apply Cancel ?

You can look at the predictions by right-clicking on the predictor node and then selecting *Classified Data*.

**Classified Data - 3:5 - Decision Tree Predictor**

File Edit HiLite Navigation View

Table "default" - Rows: 45 Spec - Columns: 6 Properties Flow Variables

Row ID	D sepal le...	D sepal w...	D petal le...	D petal wi...	S class	S Predict...
Row0	5.1	3.5	1.4	0.2	Iris-setosa	Iris-setosa
Row3	4.6	3.1	1.5	0.2	Iris-setosa	Iris-setosa
Row5	5.4	3.9	1.7	0.4	Iris-setosa	Iris-setosa
Row9	4.9	3.1	1.5	0.1	Iris-setosa	Iris-setosa
Row20	5.4	3.4	1.7	0.2	Iris-setosa	Iris-setosa
Row23	5.1	3.3	1.7	0.5	Iris-setosa	Iris-setosa
Row24	4.8	3.4	1.9	0.2	Iris-setosa	Iris-setosa
Row26	5	3.4	1.6	0.4	Iris-setosa	Iris-setosa
Row33	5.5	4.2	1.4	0.2	Iris-setosa	Iris-setosa
Row35	5	3.2	1.2	0.2	Iris-setosa	Iris-setosa
Row36	5.5	3.5	1.3	0.2	Iris-setosa	Iris-setosa
Row38	4.4	3	1.3	0.2	Iris-setosa	Iris-setosa
Row47	4.6	3.2	1.4	0.2	Iris-setosa	Iris-setosa
Row50	7	3.2	4.7	1.4	Iris-versicolor	Iris-versicolor
Row53	5.5	2.3	4	1.3	Iris-versicolor	Iris-versicolor
Row57	4.9	2.4	3.3	1	Iris-versicolor	Iris-versicolor
Row59	5.2	2.7	3.9	1.4	Iris-versicolor	Iris-versicolor
Row61	5.9	3	4.2	1.5	Iris-versicolor	Iris-versicolor
Row67	5.8	2.7	4.1	1	Iris-versicolor	Iris-versicolor
Row68	6.2	2.2	4.5	1.5	Iris-versicolor	Iris-versicolor
Row73	6.1	2.8	4.7	1.2	Iris-versicolor	Iris-versicolor
Row81	5.5	2.4	3.7	1	Iris-versicolor	Iris-versicolor
Row82	5.8	2.7	3.9	1.2	Iris-versicolor	Iris-versicolor
Row83	6	2.7	5.1	1.6	Iris-versicolor	Iris-versicolor
Row85	6	3.4	4.5	1.6	Iris-versicolor	Iris-versicolor
Row87	6.3	2.3	4.4	1.3	Iris-versicolor	Iris-versicolor
Row91	6.1	3	4.6	1.4	Iris-versicolor	Iris-versicolor
Row95	5.7	3	4.2	1.2	Iris-versicolor	Iris-versicolor
Row96	5.7	2.9	4.2	1.3	Iris-versicolor	Iris-versicolor
Row99	5.7	2.8	4.1	1.3	Iris-versicolor	Iris-versicolor
Row101	5.8	2.7	5.1	1.9	Iris-virginica	Iris-virginica
Row103	6.3	2.9	5.6	1.8	Iris-virginica	Iris-virginica
Row104	6.5	3	5.8	2.2	Iris-virginica	Iris-virginica
Row105	7.6	3	6.6	2.1	Iris-virginica	Iris-virginica
Row110	6.5	3.2	5.1	2	Iris-virginica	Iris-virginica
Row112	6.8	3	5.5	2.1	Iris-virginica	Iris-virginica
Row114	5.8	2.8	5.1	2.4	Iris-virginica	Iris-virginica
Row119	6	2.2	5	1.5	Iris-virginica	Iris-versicolor

8. Lastly, to calculate the accuracy statistics, connect the **Scorer** node with the predictor node and configure it.

Dialog - 3:6 - Scorer

File

Scorer | Flow Variables | Memory Policy

First Column  
S | class

Second Column  
S | Prediction (class)

Sorting of values in tables  
Sorting strategy: Insertion order ☐ Reverse order

Provide scores as flow variables  
☐ Use name prefix

Missing values  
In case of missing values: ☒ Ignore ☐ Fail

OK Apply Cancel ?

Once the configuration is done, execute the node and select “View: Confusion Matrix” to get the accuracy rate of the model.

Confusion Matrix - 3:6 - Scorer

File Hilite

class \ Pre...	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	13	0	0
Iris-versicolor	0	16	1
Iris-virginica	0	1	14

Correct classified: 43      Wrong classified: 2  
Accuracy: 95.556 %      Error: 4.444 %  
Cohen's kappa (κ) 0.933

9. Now, go back the learner node and select the pruning method. Then execute the model again and observe the accuracy statistics. Is there any changes?

Dialog - 3:4 - Decision Tree Learner

File

Options PMMLSettings Flow Variables

General

Class column **S** class

Quality measure Gini index

Pruning method **MDL**

☒ Reduced Error Pruning

Min number records per node 2

Number records to store for view 10,000

☒ Average split point

Number threads 4

☒ Skip nominal columns without domain information

Root split

☐ Force root split column

Root split column **D** petal width

Binary nominal splits

☐ Binary nominal splits

Max #nominal 10

☐ Filter invalid attribute values in child nodes

OK Apply Cancel ?

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow with the following nodes: File Reader (Node 1), Column Rename (Node 2), Partitioning (Node 3), Decision Tree Learner (Node 4), Decision Tree Predictor (Node 5), and Scorer (Node 6). The left sidebar contains the 'KNIME Explorer' with a tree view of the project structure, including folders for 'My KNIME Hub', 'EXAMPLES', and 'LOCAL (Local Workspace)'. Below this is the 'Workflow Coach' and the 'Node Repository' which lists various nodes like 'File Reader', 'Column Rename', 'Partitioning', 'Decision Tree Learner', 'Decision Tree Predictor', and 'Scorer'. The right sidebar shows the 'Description' tab for the current workflow, which is titled 'Week 5'. The bottom of the interface features a 'Console' window displaying log messages and a 'Node Monitor' window showing a small diagram of the workflow.