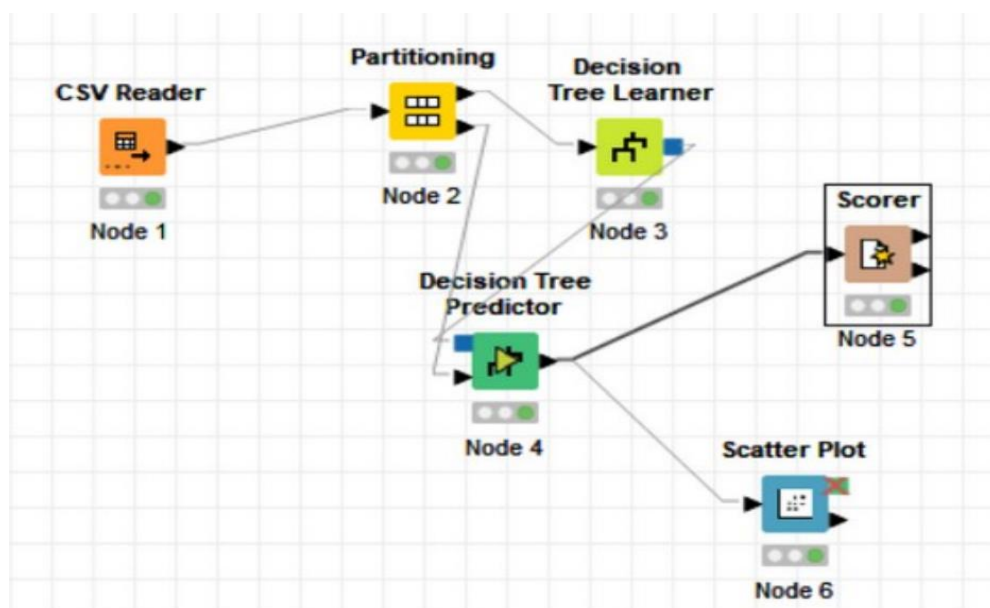


## 1. CLASSIFICATION (DECISION TABLE) WORKFLOW



Initially,

Dialog - 3:3 - Decision Tree Learner

— □ ×

File

Options PMMLSettings Flow Variables

General

Class column **S** class

Quality measure Gini index

Pruning method No pruning

☒ Reduced Error Pruning

Min number records per node 2

Number records to store for view 10,000

☒ Average split point

Number threads 8

☒ 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 ?

## Confusion matrix - 3:5 - Scorer

File Edit Hilite Navigation View

Table "spec\_name" - Rows: 3 Spec - Columns: 3 Properties Flow Variables

Row ID	Iris-virg...	Iris-ver...	Iris-set...
Iris-virginica	17	0	0
Iris-versicolor	3	10	0
Iris-setosa	0	0	15

## Accuracy statistics - 3:5 - Scorer

File Edit Hilite Navigation View

Table "default" - Rows: 4 Spec - Columns: 11 Properties Flow Variables

Row ID	I TruePo...	I FalsePo...	I TrueNe...	I FalseN...	D Recall	D Precision	D Sensitivity	D Specifcity	D F-meas...	D Accuracy	D Cohen'...
Iris-virginica	17	3	25	0	1	0.85	1	0.893	0.919	?	?
Iris-versicolor	10	0	32	3	0.769	1	0.769	1	0.87	?	?
Iris-setosa	15	0	30	0	1	1	1	1	1	?	?
Overall	?	?	?	?	?	?	?	?	?	0.933	0.898

On changing minimum number of records per node , number records to store for view and number threads

## Dialog - 3:3 - 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



## Confusion matrix - 3:5 - Scorer

File Edit Hilite Navigation View

Table "spec\_name" - Rows: 3 Spec - Columns: 3 Properties Flow Variables

Row ID	Iris-set...	Iris-ver...	Iris-virg...
Iris-setosa	10	0	0
Iris-versicolor	0	19	0
Iris-virginica	0	0	16

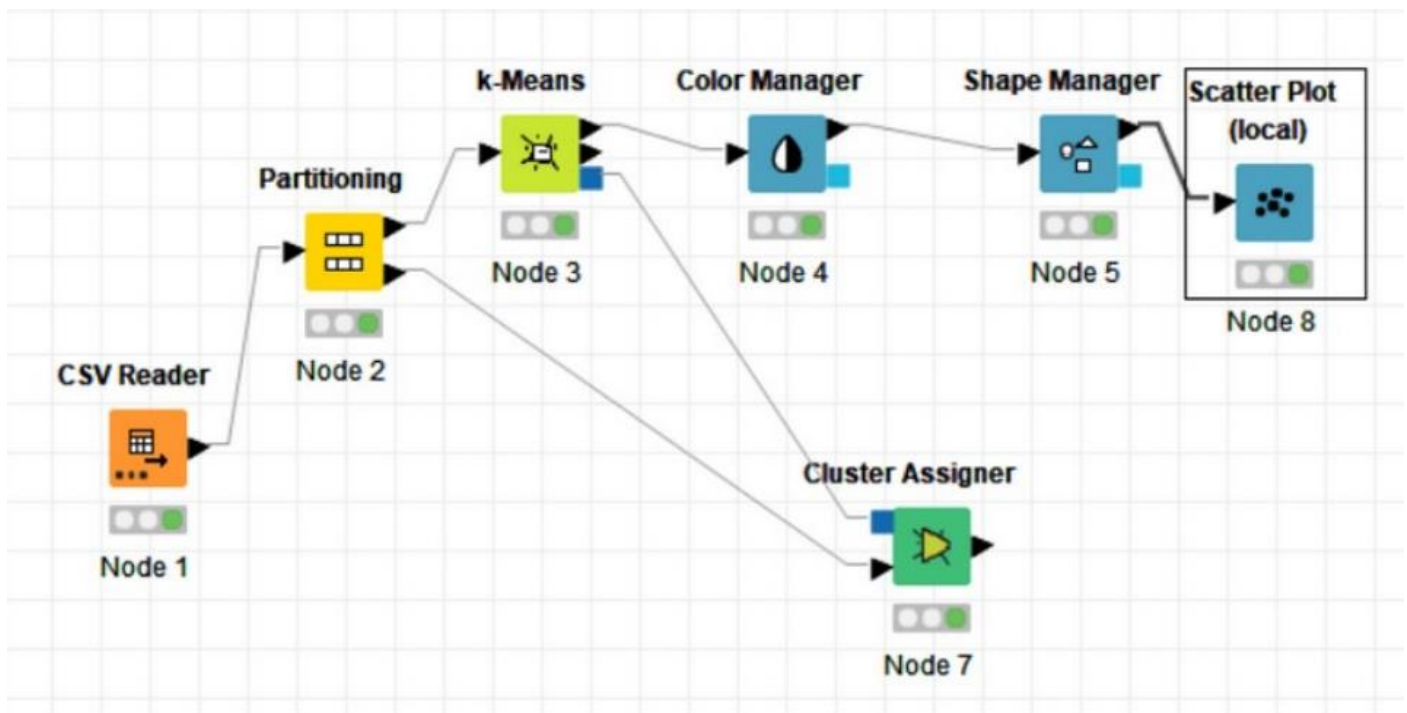
## Accuracy statistics - 3:5 - Scorer

File Edit Hilite Navigation View

Table "default" - Rows: 4 Spec - Columns: 11 Properties Flow Variables

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	D Recall	D Precision	D Sensitivity	D Specficity	D F-meas...	D Accuracy	D Cohen'...
Iris-setosa	10	0	35	0	1	1	1	1	1	?	?
Iris-versicolor	19	0	26	0	1	1	1	1	1	?	?
Iris-virginica	16	0	29	0	1	1	1	1	1	?	?
Overall	?	?	?	?	?	?	?	?	?	1	1

## 2. CLUSTERING (K MEANS) WORKFLOW



File

K-Means Properties   Flow Variables   Memory Policy

Clusters

Number of clusters:

Centroid initialization:

☐ First k rows

☒ Random initialization ☒ Use static random seed

Number of Iterations

Max. number of iterations:

Column Selection

Exclude

Filter

>

>>

<

<<

Include

Filter

☒ sepal length

☒ sepal width

☒ petal length

☒ petal width

☐ Always include all columns

Hilite Mapping

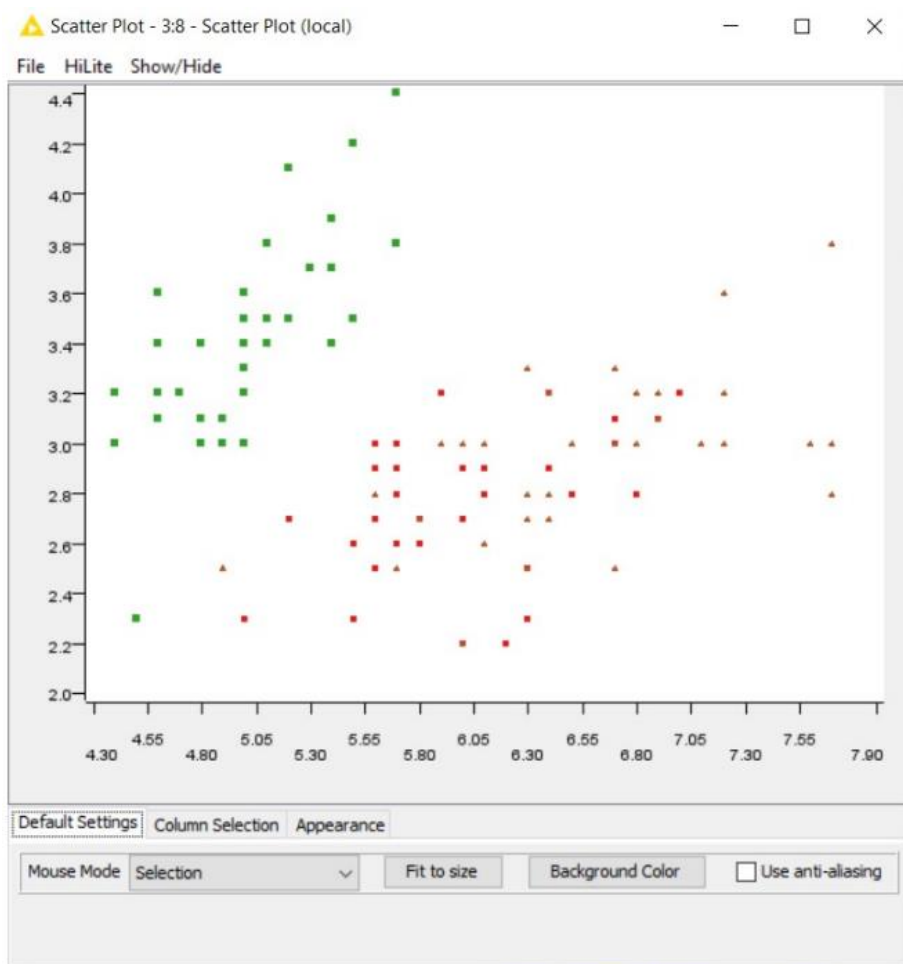
☐ Enable Hilite Mapping

## Assigned Data - 3:7 - Cluster Assigner

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 Cluster
Row7	5	3.4	1.5	0.2	Iris-setosa	cluster_0
Row8	4.4	2.9	1.4	0.2	Iris-setosa	cluster_0
Row11	4.8	3.4	1.6	0.2	Iris-setosa	cluster_0
Row13	4.3	3	1.1	0.1	Iris-setosa	cluster_0
Row14	5.8	4	1.2	0.2	Iris-setosa	cluster_0
Row21	5.1	3.7	1.5	0.4	Iris-setosa	cluster_0
Row23	5.1	3.3	1.7	0.5	Iris-setosa	cluster_0
Row28	5.2	3.4	1.4	0.2	Iris-setosa	cluster_0
Row31	5.4	3.4	1.5	0.4	Iris-setosa	cluster_0
Row44	5.1	3.8	1.9	0.4	Iris-setosa	cluster_0
Row56	6.3	3.3	4.7	1.6	Iris-versicolor	cluster_1
Row57	4.9	2.4	3.3	1	Iris-versicolor	cluster_1
Row58	6.6	2.9	4.6	1.3	Iris-versicolor	cluster_1
Row60	5	2	3.5	1	Iris-versicolor	cluster_1
Row61	5.9	3	4.2	1.5	Iris-versicolor	cluster_1
Row71	6.1	2.8	4	1.3	Iris-versicolor	cluster_1
Row75	6.6	3	4.4	1.4	Iris-versicolor	cluster_1
Row80	5.5	2.4	3.8	1.1	Iris-versicolor	cluster_1
Row81	5.5	2.4	3.7	1	Iris-versicolor	cluster_1
Row82	5.8	2.7	3.9	1.2	Iris-versicolor	cluster_1
Row84	5.4	3	4.5	1.5	Iris-versicolor	cluster_1
Row85	6	3.4	4.5	1.6	Iris-versicolor	cluster_1
Row86	6.7	3.1	4.7	1.5	Iris-versicolor	cluster_1
Row88	5.6	3	4.1	1.3	Iris-versicolor	cluster_1
Row89	5.5	2.5	4	1.3	Iris-versicolor	cluster_1
Row91	6.1	3	4.6	1.4	Iris-versicolor	cluster_1
Row97	6.2	2.9	4.3	1.3	Iris-versicolor	cluster_1
Row98	5.1	2.5	3	1.1	Iris-versicolor	cluster_1
Row99	5.7	2.8	4.1	1.3	Iris-versicolor	cluster_1
Row103	6.3	2.9	5.6	1.8	Iris-virginica	cluster_2
Row104	6.5	3	5.8	2.2	Iris-virginica	cluster_2
Row107	7.3	2.9	6.3	1.8	Iris-virginica	cluster_2
Row110	6.5	3.2	5.1	2	Iris-virginica	cluster_2
Row114	5.8	2.8	5.1	2.4	Iris-virginica	cluster_1
Row116	6.5	3	5.5	1.8	Iris-virginica	cluster_2
Row118	7.7	2.6	6.9	2.3	Iris-virginica	cluster_2
Row126	6.2	2.8	4.8	1.8	Iris-virginica	cluster_1
Row128	6.4	2.8	5.6	2.1	Iris-virginica	cluster_2



After modifying number of clusters to five

Dialog - 3:3 - k-Means

File

K-Means Properties Flow Variables Memory Policy

Clusters

Number of clusters:

Centroid initialization:

☐ First k rows

☒ Random initialization ☒ Use static random seed

Number of Iterations

Max. number of iterations:

Column Selection

Exclude

Filter

> >> < <<

Include

Filter

☒ Always include all columns

HiLite Mapping

☐ Enable HiLite Mapping

OK Apply Cancel ?



# Assigned Data - 3:7 - Cluster Assigner

File Edit Hilite Navigation View

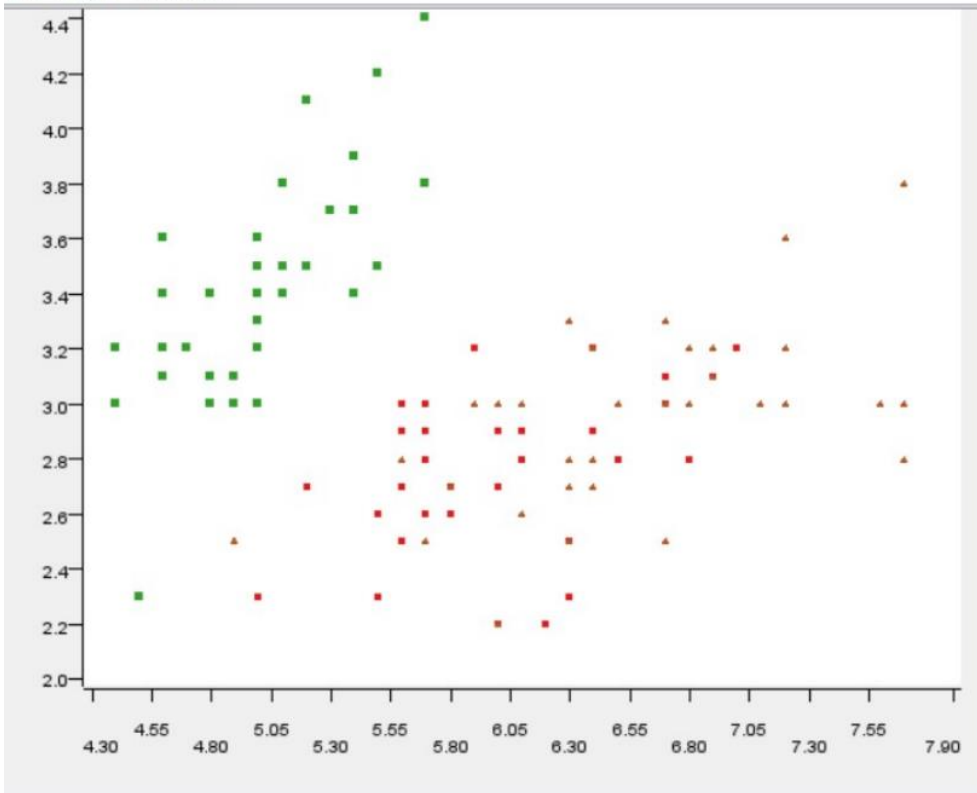
Table "default" - Rows: 45 Spec - Columns: 6 Properties Flow Variables

Row ID	D sepal le...	D sepal w...	D petal le...	D petal wi...	S class	S Cluster
Row7	5	3.4	1.5	0.2	Iris-setosa	cluster_2
Row8	4.4	2.9	1.4	0.2	Iris-setosa	cluster_1
Row11	4.8	3.4	1.6	0.2	Iris-setosa	cluster_2
Row13	4.3	3	1.1	0.1	Iris-setosa	cluster_1
Row14	5.8	4	1.2	0.2	Iris-setosa	cluster_0
Row21	5.1	3.7	1.5	0.4	Iris-setosa	cluster_2
Row23	5.1	3.3	1.7	0.5	Iris-setosa	cluster_2
Row28	5.2	3.4	1.4	0.2	Iris-setosa	cluster_2
Row31	5.4	3.4	1.5	0.4	Iris-setosa	cluster_2
Row44	5.1	3.8	1.9	0.4	Iris-setosa	cluster_0
Row56	6.3	3.3	4.7	1.6	Iris-versicolor	cluster_3
Row57	4.9	2.4	3.3	1	Iris-versicolor	cluster_3
Row58	6.6	2.9	4.6	1.3	Iris-versicolor	cluster_3
Row60	5	2	3.5	1	Iris-versicolor	cluster_3
Row61	5.9	3	4.2	1.5	Iris-versicolor	cluster_3
Row71	6.1	2.8	4	1.3	Iris-versicolor	cluster_3
Row75	6.6	3	4.4	1.4	Iris-versicolor	cluster_3
Row80	5.5	2.4	3.8	1.1	Iris-versicolor	cluster_3
Row81	5.5	2.4	3.7	1	Iris-versicolor	cluster_3
Row82	5.8	2.7	3.9	1.2	Iris-versicolor	cluster_3
Row84	5.4	3	4.5	1.5	Iris-versicolor	cluster_3
Row85	6	3.4	4.5	1.6	Iris-versicolor	cluster_3
Row86	6.7	3.1	4.7	1.5	Iris-versicolor	cluster_3
Row88	5.6	3	4.1	1.3	Iris-versicolor	cluster_3
Row89	5.5	2.5	4	1.3	Iris-versicolor	cluster_3
Row91	6.1	3	4.6	1.4	Iris-versicolor	cluster_3
Row97	6.2	2.9	4.3	1.3	Iris-versicolor	cluster_3
Row98	5.1	2.5	3	1.1	Iris-versicolor	cluster_3
Row99	5.7	2.8	4.1	1.3	Iris-versicolor	cluster_3
Row103	6.3	2.9	5.6	1.8	Iris-virginica	cluster_4
Row104	6.5	3	5.8	2.2	Iris-virginica	cluster_4
Row107	7.3	2.9	6.3	1.8	Iris-virginica	cluster_4
Row110	6.5	3.2	5.1	2	Iris-virginica	cluster_4
Row114	5.8	2.8	5.1	2.4	Iris-virginica	cluster_3
Row116	6.5	3	5.5	1.8	Iris-virginica	cluster_4
Row118	7.7	2.6	6.9	2.3	Iris-virginica	cluster_4
Row126	6.2	2.8	4.8	1.8	Iris-virginica	cluster_3
Row128	6.4	2.8	5.6	2.1	Iris-virginica	cluster_4

## Scatter Plot - 3:6 - Scatter Plot (local)

— □ ×

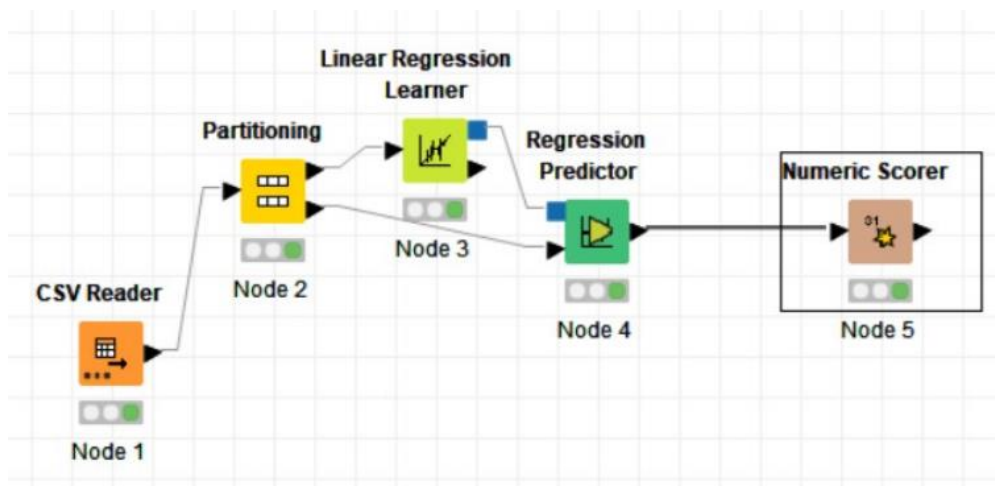
File Hilite Show/Hide



Default Settings Column Selection Appearance

Mouse Mode Selection Fit to size Background Color Use anti-aliasing

### 3. REGRESSION (LINEAR REGRESSION) WORKFLOW



Initially:

Dialog - 3:3 - Linear Regression Learner

File

Settings Flow Variables Memory Policy

Target: D petal width

Values

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

Filter

D sepal length  
D sepal width  
D petal length  
S class

☐ Enforce inclusion

Regression Properties

☐ Predefined Offset Value: 0

Missing Values in Input Data

☐ Ignore rows with missing values.  
☒ Fail on observing missing values.

Scatter Plot View

First Row: 1  
Row Count: 20,000

OK Apply Cancel ?

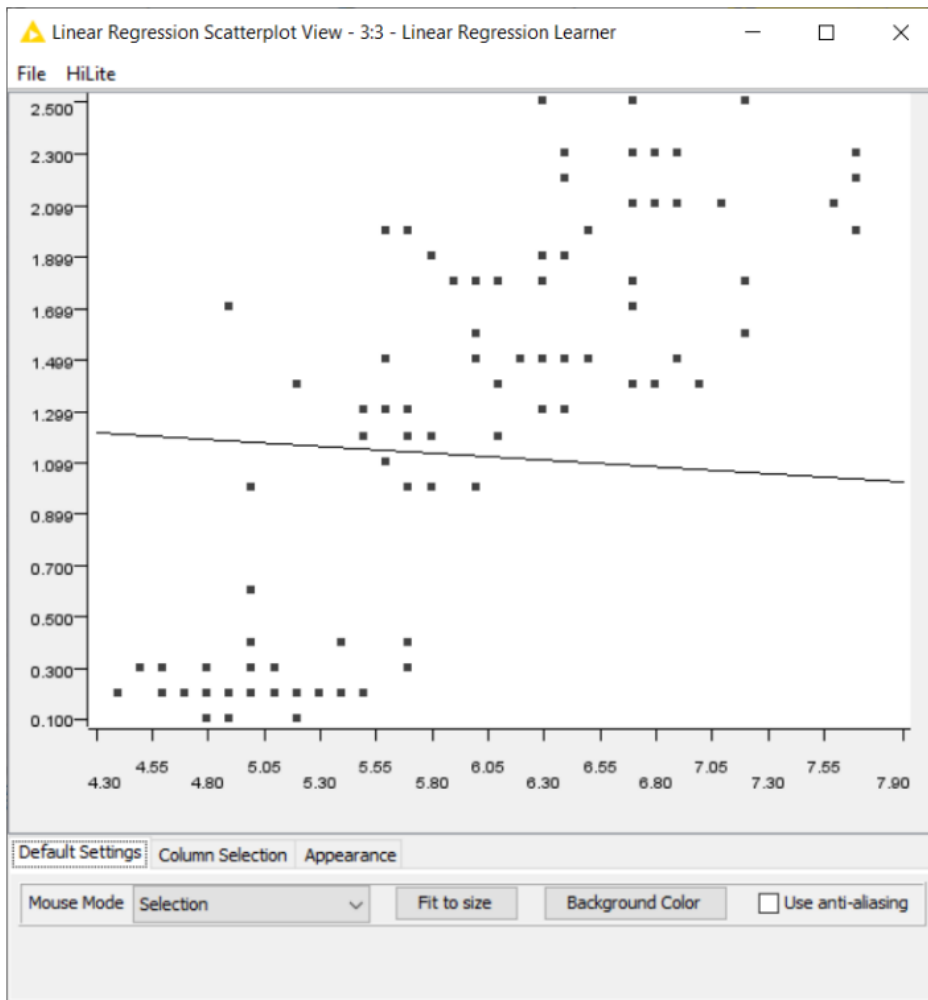
Linear Regression Result View - 3...

File

**Statistics on Linear Regression**

Variable	Coeff.	Std. Err.	t-value	P> t
sepal length	-0.0538	0.0554	-0.9702	0.3343
sepal width	0.2239	0.0582	3.8482	0.0002
petal length	0.2275	0.0615	3.6991	0.0004
class=Iris-versicolor	0.6494	0.1594	4.074	9.33E-5
class=Iris-virginica	1.0441	0.208	5.0203	2.28E-6
Intercept	-0.5929	0.2131	-2.7829	0.0065

Multiple R-Squared: 0.9549  
Adjusted R-Squared: 0.9526



## On changing Target and predefined offset value

Dialog - 3:3 - Linear Regression Learner

File

Settings Flow Variables Memory Policy

Target

D sepal width

Values

Manual Selection Wildcard/Regex Selection

Exclude

Filter

No columns in this list

Enforce exclusion

Include

Filter

D sepal length  
D petal length  
D petal width  
S class

Enforce inclusion

Regression Properties

Predefined Offset Value: 5

Missing Values in Input Data

Ignore rows with missing values.  
Fail on observing missing values.

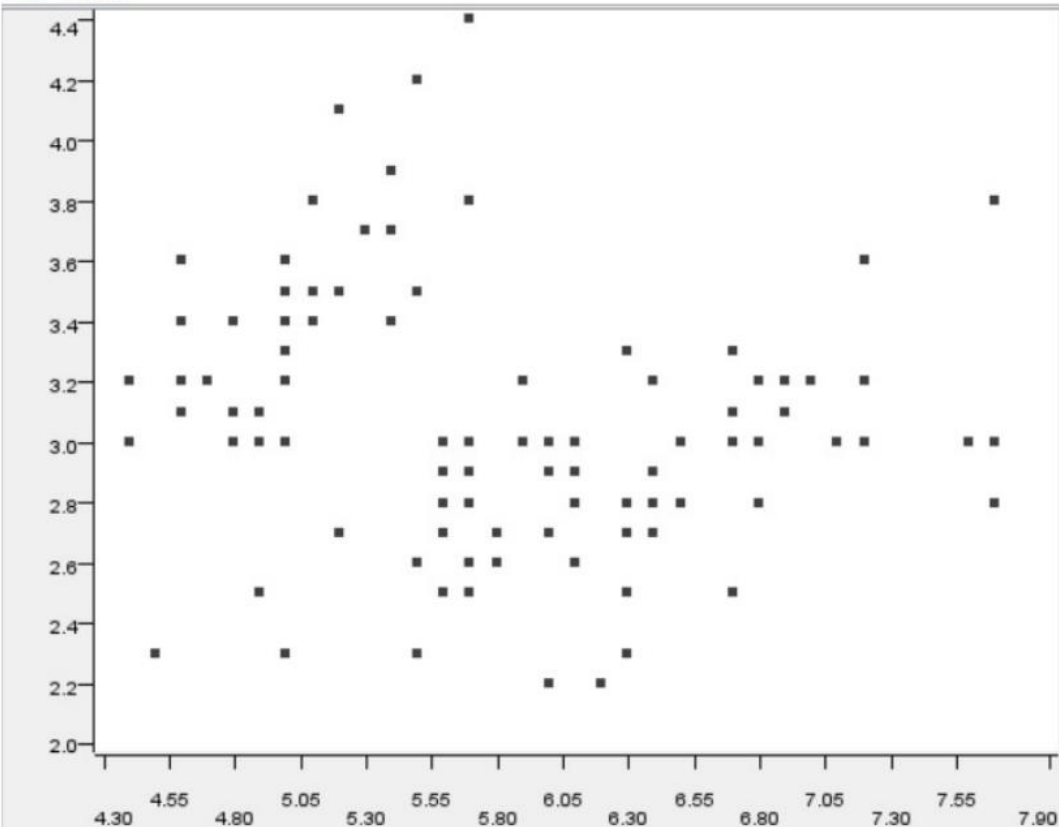
Scatter Plot View

First Row: 1  
Row Count: 20,000

OK Apply Cancel ?



File HiLite



Default Settings Column Selection Appearance

Mouse Mode

Selection



Fit to size

Background Color



Use anti-aliasing

File

## Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
sepal length	-0.4224	0.0403	-10.4831	0.0
petal length	0.2789	0.1381	2.0197	0.0461
petal width	0.7866	0.2186	3.5979	0.0005
class=Iris-versicolor	-1.9776	0.3502	-5.647	1.54E-7
class=Iris-virginica	-2.3884	0.4904	-4.8699	4.19E-6

Offset Value: 5

Multiple R-Squared: 0.9606

Adjusted R-Squared: 0.9587