

Aufgabe – Workcamp Maschinelles Lernen

Regression Tree zur Prediction von Sensordaten.

Laden Sie die Datei **IHK-Sensordaten-Workcamp-Aufgabe.knar** vom github Repository hier herunter

<https://github.com/mediagit2016/workcamp-maschinelles-lernen-grundlagen/tree/master/20-02-20-ml-workcamp/aufgaben>

Importieren Sie das .knar File in KNIME

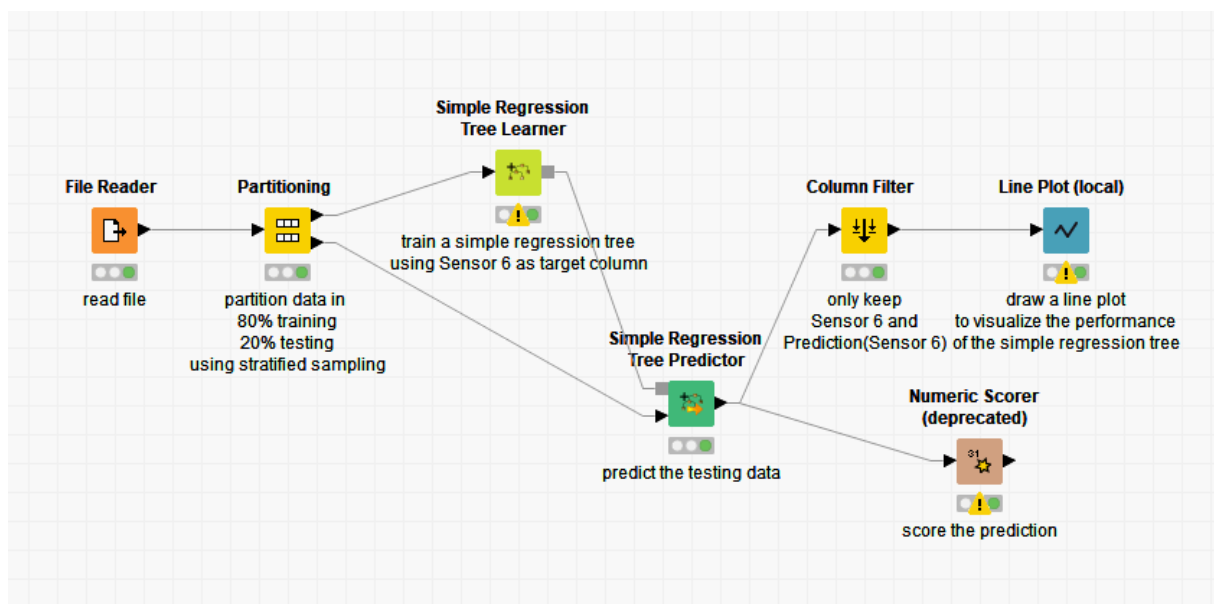
IM Workflow finden Sie bereits einen File Reader, Numeric Scorer und Plot Knoten.

Die Daten enthalten Werte der Sensoren 1,2,3,5,6 mit deren Hilfe eine Kategorisierung der Werkstücke vorgenommen wird. In einer späteren Aufgabe trainieren wir hierfür ein Machine Learning Modell.

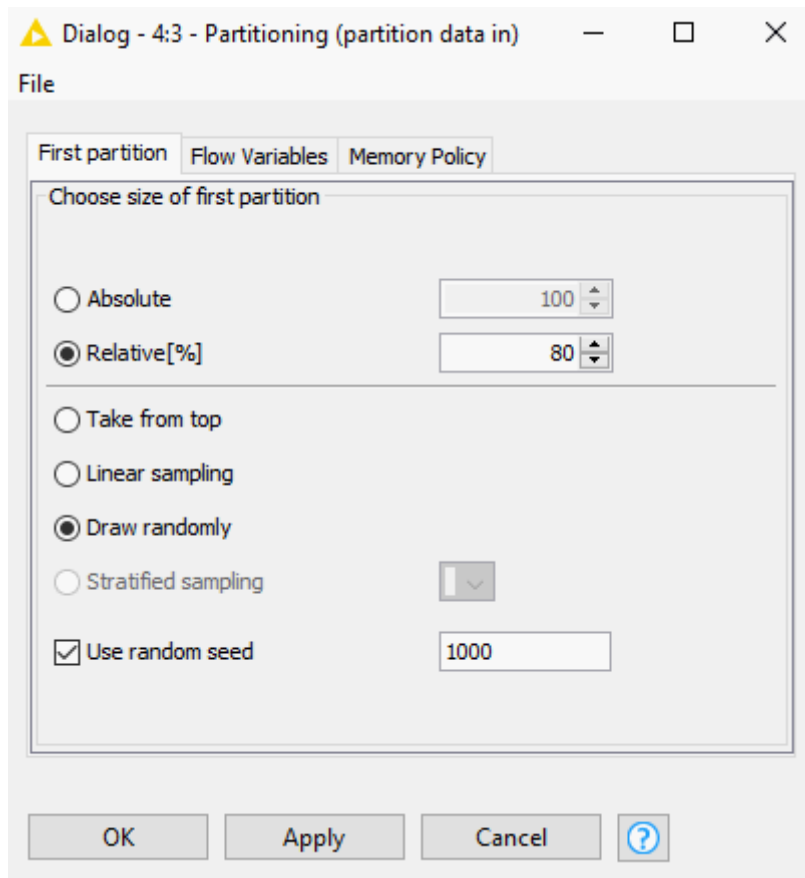
Jetzt soll mit Hilfe der Sensoren 1,2,3,5 der Wert des Sensors 6 vorhergesagt werden können.

Wir verwenden hierzu die Methode eines Regression Trees.

Aufbau:



Ziehen Sie einen Partitioning Node in den Workflow und konfigurieren den Node folgendermaßen:

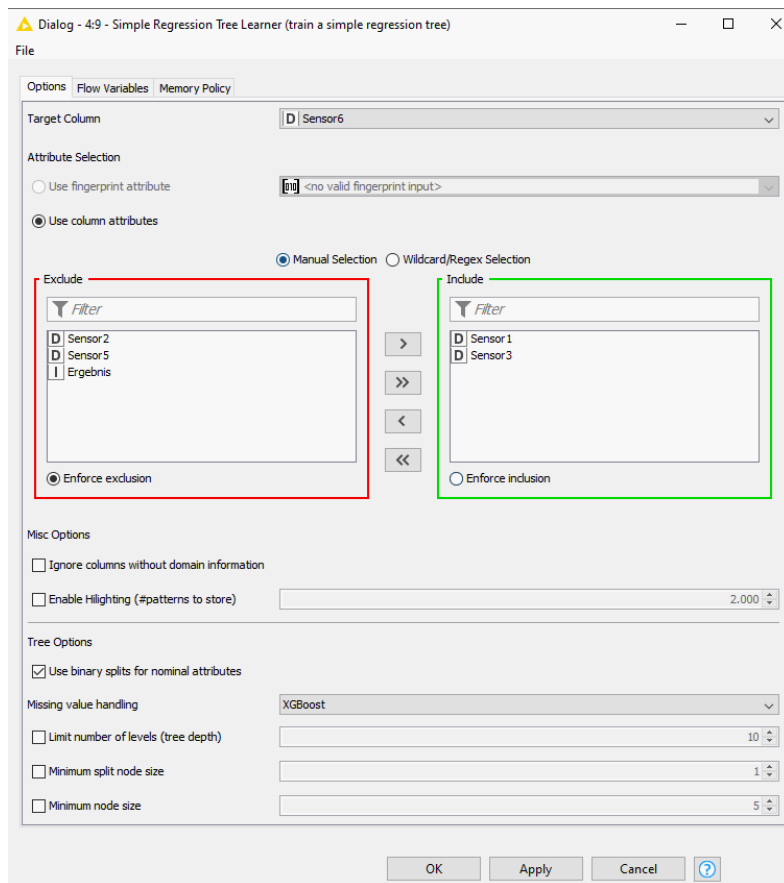


Nicht vergessen, das random seed auf 1000 zu stellen, damit wir alle gleiche Werte bekommen.

Verbinden Sie den Partitioning Node mit dem File Reader

Führen Sie den Partitioning Node aus

Ziehen Sie einen Simple Regression Tree Node auf denWorkspace



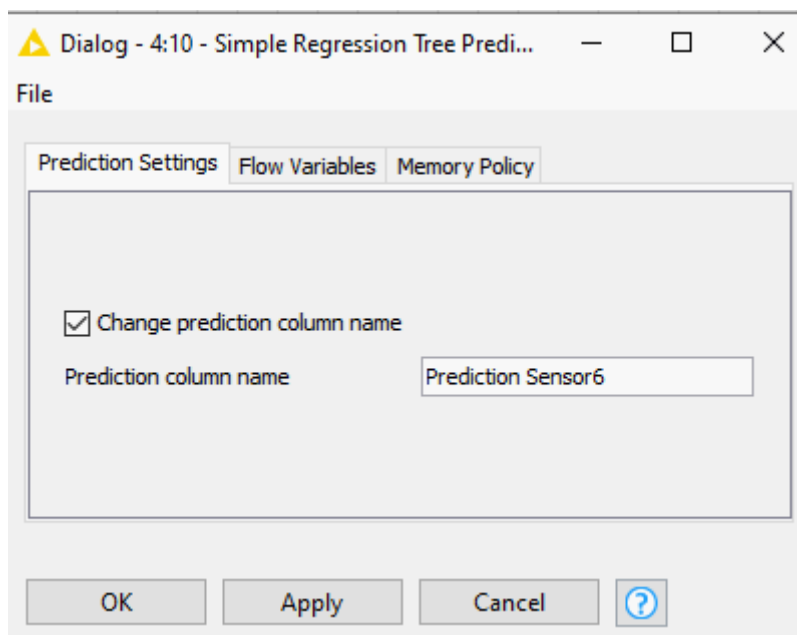
Konfigurieren Sie den Learner Node wie oben:

Target Column: Sensor 6

Im ersten Schritt wählen wir nur Sensor 1 und Sensor 3 zur Vorhersage aus.

Verbinden Sie den oberen Port des Partitioning Nodes mit dem Learner Node

Ziehen Sie einen Regression Tree Predictor Node auf den Workspace und konfigurieren wie folgt:



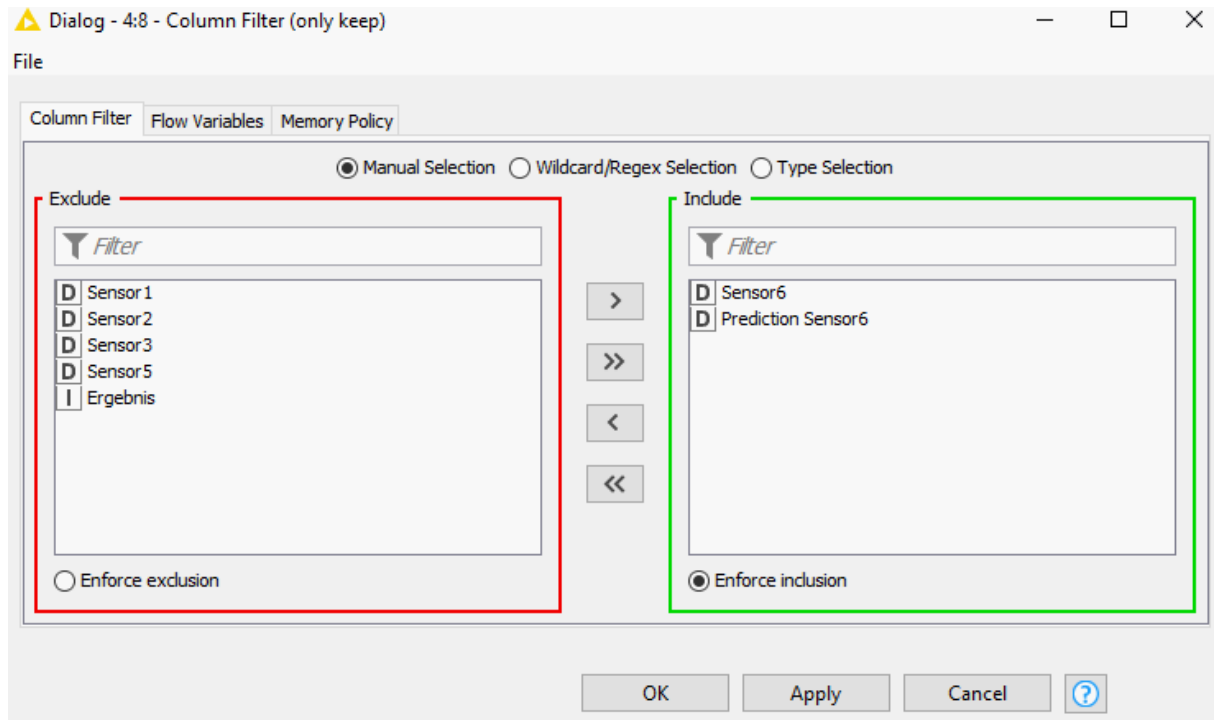
Wir erhalten dadurch eine neue Spalte: Prediction Sensor 6

Verbinden Sie den Modell Port des Learners mit dem Predictor

Verbinden Sie den unteren Port des Partitioning Nodes mit demEingang Port des Predictors.

Ziehen Sie einen Column Filter Node auf den Workspace

Verbinden Sie den Ausgangsport des Predictors mit diesem Column Filter Node und konfigurieren Sie



Verbinden Sie den Line Plot Node mit dem Ausgangsport des Column Filter Nodes

Verbinden Sie den Ausgangsport des Predictor Nodes mit dem Numeric Scorer Node und konfigurieren Sie den Scorer:

Dialog - 4:5 - Numeric Scorer (deprecated) (...)

File

Options Flow Variables Memory Policy

Reference column

Predicted column

Output column

☐ Change column name

Output column name

Provide scores as flow variables

Prefix of flow variables

☐ Output scores as flow variables

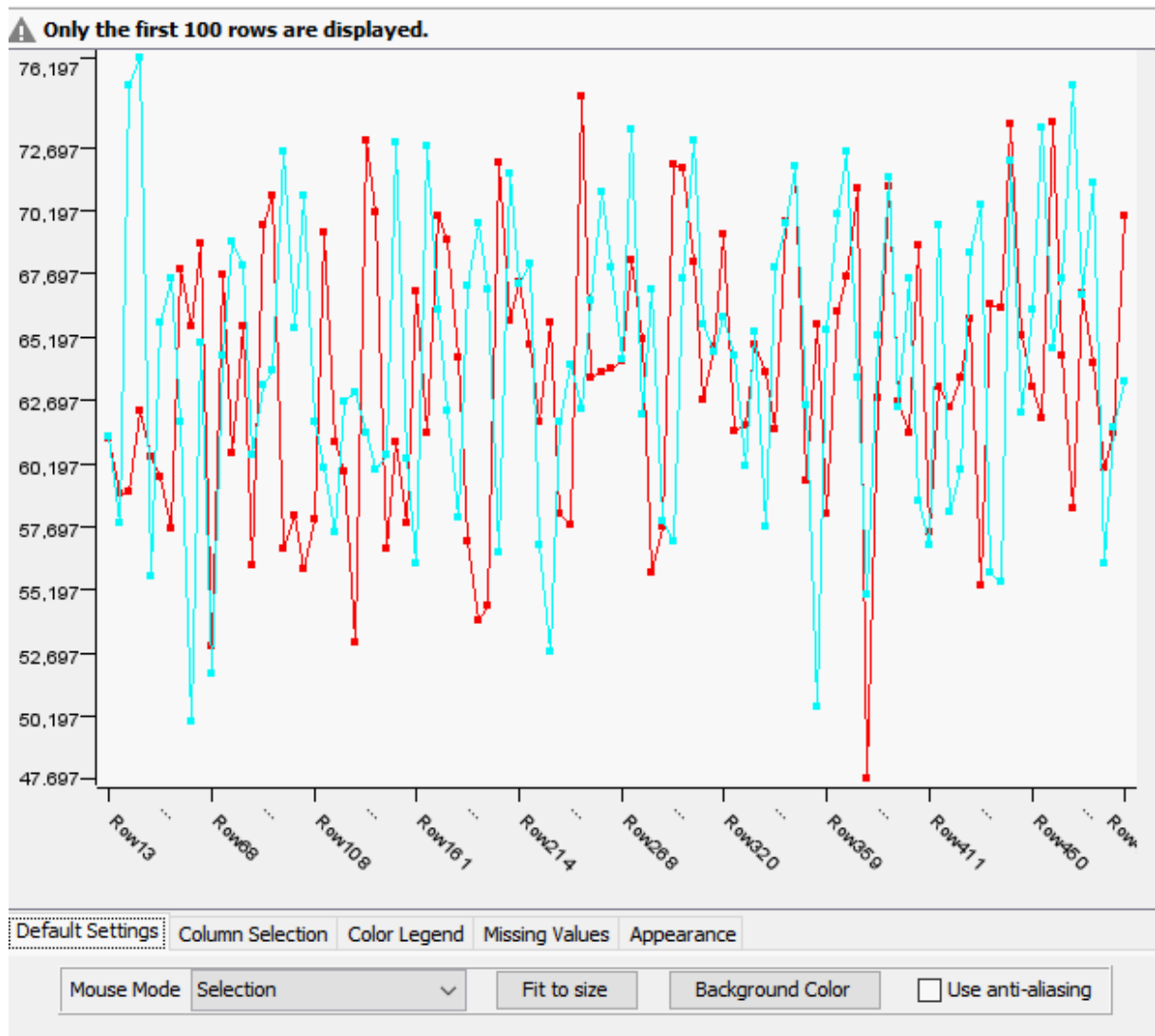
OK Apply Cancel ?

ReferenceColumn: Sensor 6

Predicted Column: Prediction Sensor 6

Wenn alle Knoten richtig verbunden und Konfiguriert sind, führen Sie den gesamten Workflow aus.

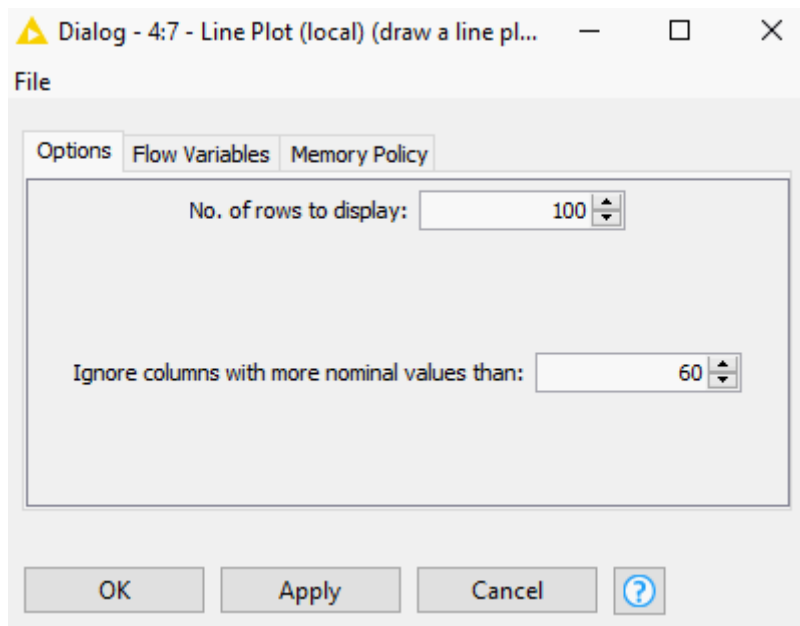
Im Plot Line Node erhalten Sie folgendes Bild:



Wir erhalten eine relativ schlechte Vorhersage der Werte.

Rot: Daten des Sensors 6, Blau Prediction des Sensors 6

Sie müssen dazu den Plot Node konfigurieren:



Num of rows to display auf 100 setzen.

Mit View Plot sehen Sie obige Abbildung.

Im Numeric Scorer erhalten Sie folgende Statistik:

Statistics - 4:5 - Numeric Scorer

File Hilite Navigation View

Table "Scores" - Rows: 5 Spec - Column

Row ID	Predict...
R ²	-0.918
mean absolut...	6.486
mean square...	67.178
root mean sq...	8.196
mean signed ...	-0.144

Einen sehr schlechten R² Wert

Ergebnis:

Mit Sensor 1 und 3 lassen sich die Sensordaten für Sensor 6 nur sehr ungenau prognostizieren.

Möglichkeit:

Veränderung des Regression Tree Learners:

Dialog - 4:9 - Simple Regression Tree Learner (train a simple regression tree)

File

Options | Flow Variables | Memory Policy

Target Column: D | Sensor6

Attribute Selection

☐ Use fingerprint attribute: [no valid fingerprint input]

☒ Use column attributes

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Filter

1 Ergebnis

☒ Enforce exclusion

Include

Filter

D | Sensor 1
D | Sensor 2
D | Sensor 3
D | Sensor 5

☐ Enforce inclusion

Misc Options

☐ Ignore columns without domain information

☐ Enable Highlighting (#patterns to store): 2.000

Tree Options

☒ Use binary splits for nominal attributes

Missing value handling: XGBoost

☐ Limit number of levels (tree depth): 10

☐ Minimum split node size: 1

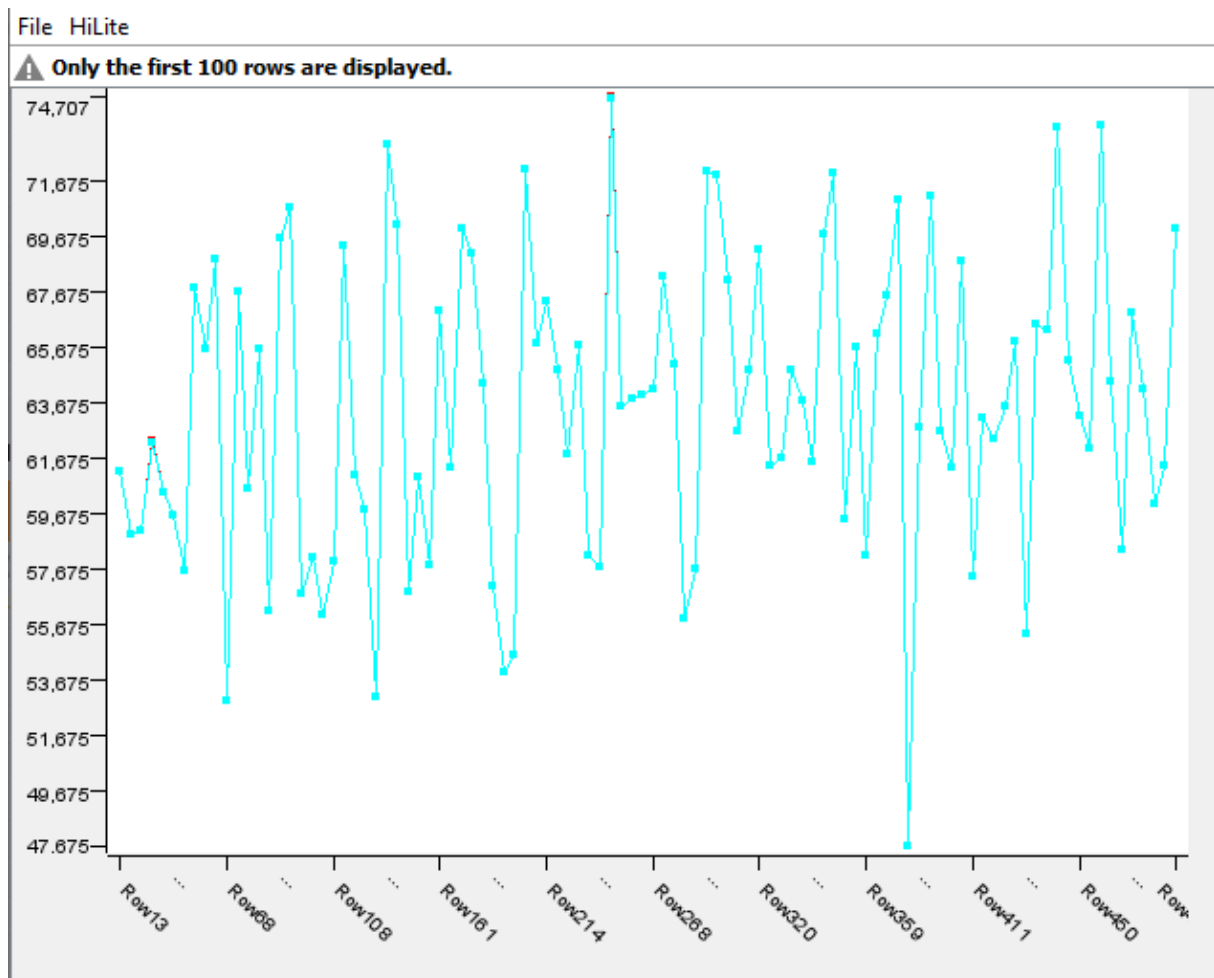
☐ Minimum node size: 5

OK Apply Cancel ?

Wir übernehmen zusätzlich Sensor 2 und 5 in den Learner.

Führen Sie jetzt den Workflow neu aus.

Im Plot erhalten wir:



Prediction Sensor6 und Sensor6 stimmen nahezu überein.

Im Scorer erhalten wir:

⚠ Statistics - 4:5 - Numeric Sco

File HiLite Navigation View

Table "Scores" - Rows: 5 Spec - Co

Row ID	Predict...
R ²	1
mean absolut...	0.003
mean square...	0.003
root mean sq...	0.053
mean signed ...	0.001

Welchen Schluss kann man daraus ziehen ?