



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Název projektu	Rozvoj vzdělávání na Slezské univerzitě v Opavě
Registrační číslo projektu	CZ.02.2.69/0.0./0.0/16_015/0002400

Dolování dat

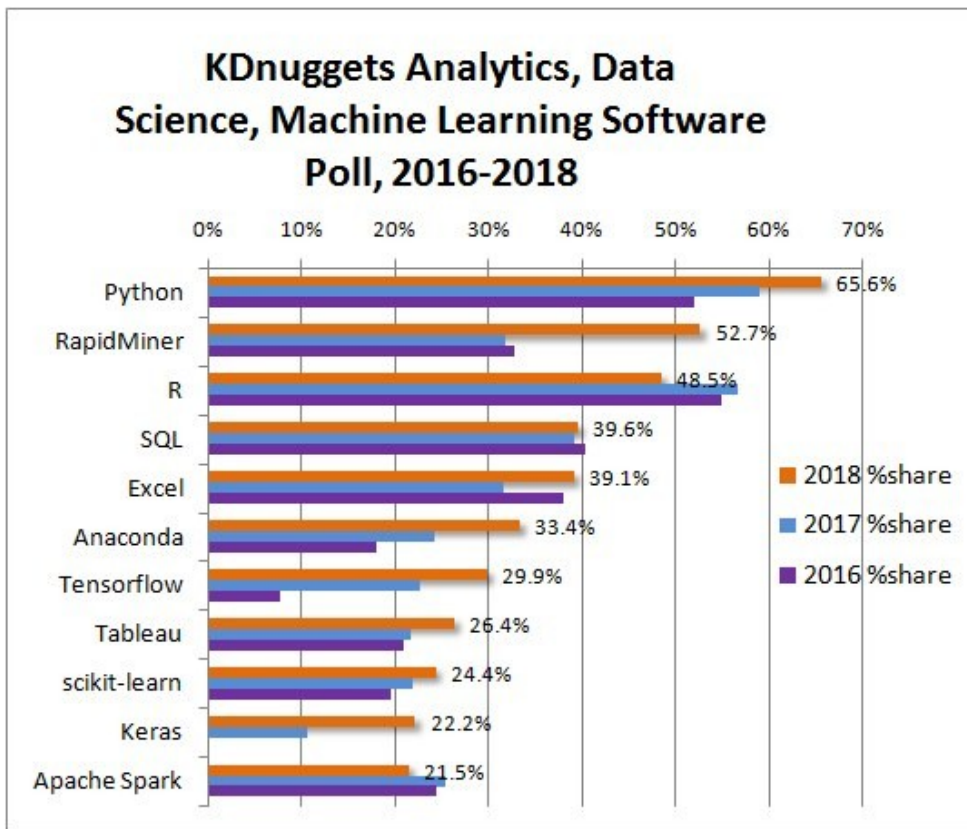
System RapidMiner

Jan Górecki



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Popularita DM systémů (2015-2017)



Educational License Program

Free Licenses for Academic Usage

<https://rapidminer.com/educational-program/> popř.

<http://rapidminer.findmysoft.com/>

	EDUCATIONAL	FREE
	Free	Free
# Data Rows	Unlimited	10,000
# Logical Processors	1	1
Auto Model	✓	
License Expiration	After 12 Months	Never
Customer Support	Community	Community
	Apply for Access	Download

Startujeme RapidMiner Studio




**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ




Začátky jsou těžké... ale s RapidMinerem rychle za námi

Choose a template to start from:


Blank
Start with a blank process.




Churn Modeling
Predict which of your customers will churn and why with a decision tree.




Direct Marketing
Predict response to campaigns and increase the conversion rate of your campaign.




Credit Risk Modeling
Model credit default risk by training an optimized Support Vector Machine (SVM) model.




Market Basket Analysis
Find products frequently purchased together and turn them into rules for recommendations.




Predictive Maintenance
Model equipment failures to schedule maintenance pre-emptively.



Price Risk Clustering
Cluster price developments using X-Means to unveil price-risk-relationships.



Lift Chart
Create a lift chart to visualize the improvement that a model provides compared to guessing.



- **Můžeme začít rovnou s předpřipravenými daty a modely**

Hlavní obrazovka – Design workspace



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

<new process> – RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: **Design** Results Auto Model

Find data, operators...etc All Studio Search

Repository

- Add Data
- Samples
- DB
- Local Repository (gor0001)
- Cloud Repository (disconnected)

Process

Process

100%

Your process looks empty.
Add some data first.
Drag data or operators here.

Parameters

- logverbosity: init
- logfile: [Folder icon]
- Show advanced parameters
- Change compatibility (8.1.001)

Operators

Search for Operators

- Data Access (47)
- Blending (77)
- Cleansing (26)
- Modeling (130)
- Scoring (12)
- Validation (29)
- Utility (85)
- Extensions (17)

Get more operators from the Marketplace

Recommended Operators

- Retrieve (12%)
- Select Attributes (6%)
- Set Role (5%)
- Apply Model (4%)

Help

Process
RapidMiner Studio Core

Synopsis
The root operator which is the outer most operator of every process.

Description
Each process must contain exactly one operator of this class, and it must be the root operator of the

Import dat



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Local Repository/moje* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model Find data, operators, etc. All Studio Search

Repository

- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - moje (gor0001 - v1, 9/22/17 11:39 AM - 3 kB)

Process

Process

100%

Parameters

Read Excel

Import Configuration Wizard...

excel file

sheet number 1

first row as names

date format Enter value...

Show advanced parameters

Change compatibility (8.1.001)

Help

Read Excel

RapidMiner Studio Core

Tags: Load, Import, Read, Data, Files, Xls, Xlsx, Microsoft, Spreadsheets, Datasets

Synopsis

This operator reads an ExampleSet from the specified Excel file.

[Jump to Tutorial Process](#)

Operators

read

- Data Access (22)
 - Files (16)
 - Read (15)
 - Read CSV
 - Read Excel**
 - Read Excel with Format
 - Read URL
 - Read SPSS
 - Read Stata

We found "Spreadsheet Table Extraction", "MLWizard" and 2 more results in the Marketplace. [Show me!](#)

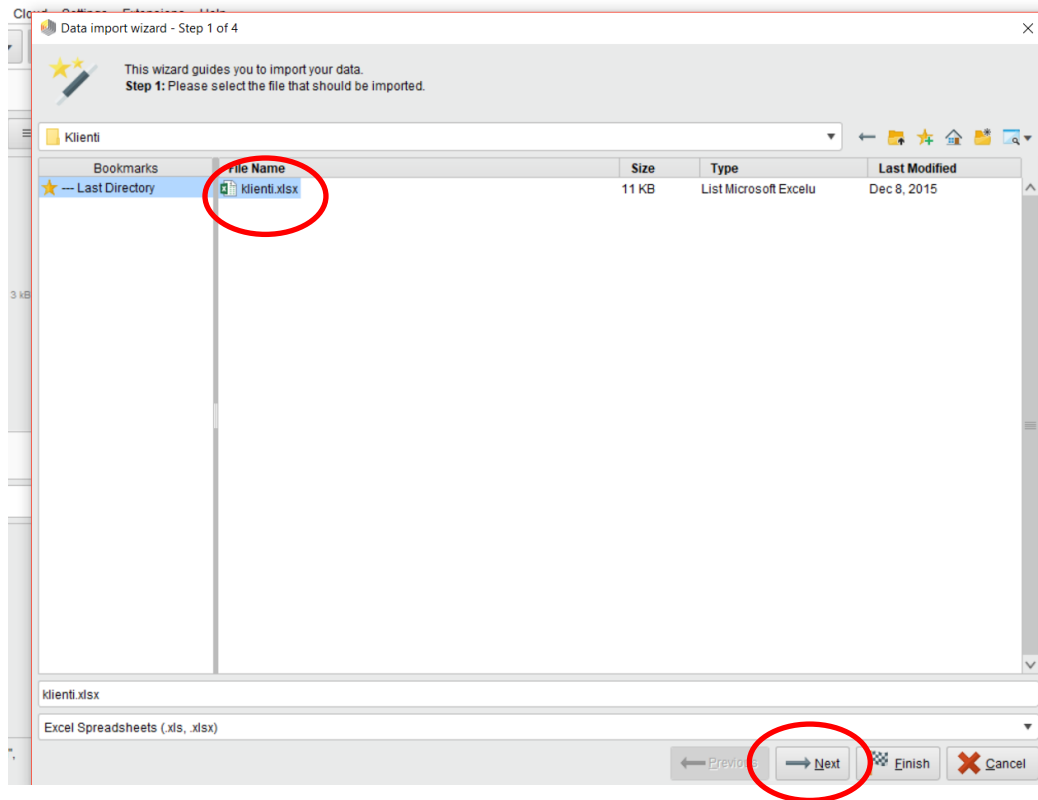
Recommended Operators

- Select Attributes 29%
- Set Role 26%
- Filter Examples 22%
- Apply Model 20%

Průvodce importem dat z Excelu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ



Průvodce importem dat z Excelu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Data import wizard - Step 2 of 4

This wizard guides you to import your data.
Step 2: An Excel file can contain multiple sheets. Please select the one you want to import into RapidMiner Studio. Furthermore, you can mark a range of cells to be loaded.

nominální numerické log2

A	B	C	D	E
příjem	konto	pohlaví	nezamě...	úvěr
vysoký	vysoké	žena	ne	ano
vysoký	vysoké	muž	ne	ano
nízký	nízké	muž	ne	ne
nízký	vysoké	žena	ano	ano
nízký	vysoké	muž	ano	ano
nízký	nízké	žena	ano	ne
vysoký	nízké	muž	ne	ano
vysoký	nízké	žena	ano	ano
nízký	střední	muž	ano	ne
vysoký	střední	žena	ne	ano
nízký	střední	žena	ano	ne
nízký	střední	muž	ne	ano

← Previous **Next** → Finish Cancel

Průvodce importem dat z Excelu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Data import wizard - Step 3 of 4

This wizard guides you to import your data.
Step 3: In RapidMiner Studio, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.

Annotat...	A	B	C	D	E
Name	příjem	konto	pohlaví	nezamé...	úvěr
-	vysoký	vysoké	žena	ne	ano
-	vysoký	vysoké	muž	ne	ano
-	nízký	nízké	muž	ne	ne
-	nízký	vysoké	žena	ano	ano
-	nízký	vysoké	muž	ano	ano
-	nízký	nízké	žena	ano	ne
-	vysoký	nízké	muž	ne	ano
-	vysoký	nízké	žena	ano	ano
-	nízký	střední	muž	ano	ne
-	vysoký	střední	žena	ne	ano
-	nízký	střední	žena	ano	ne
-	nízký	střední	muž	ne	ano

← Previous **Next** → Finish Cancel

Průvodce importem dat z Excelu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Data import wizard - Step 4 of 4

This wizard guides you to import your data.
Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely.

Date format

Preview uses only first 100 rows.

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
příjem	konto	pohlaví	nezaměstná	úvěr
polyno...	polyno...	binomi...	polyno...	polyno...
attribute	attribute	attribute	attribute	attribute
vysoký	vysoké	žena	ne	ano
vysoký	vysoké	muž	ne	ano
nízký	nízké	muž	ne	ne
nízký	vysoké	žena	ano	ano
nízký	vysoké	muž	ano	ano
nízký	nízké	žena	ano	ne
vysoký	nízké	muž	ne	ano
vysoký	nízké	žena	ano	ano
nízký	střední	muž	ano	ne
vysoký	střední	žena	ne	ano
nízký	střední	žena	ano	ne

0 errors. Ignore errors Show only errors

Row, Column	Error	Original value	Message
-------------	-------	----------------	---------

Vizualizace dat



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

The screenshot displays the RapidMiner Studio interface. The main workspace shows a process diagram with a 'Read Excel' operator highlighted by a red oval. The interface includes several panels:

- Repository:** Shows a tree view of data sources, including 'Local Repository' and 'Cloud Repository'.
- Operators:** A search bar with 'read' entered, showing a list of operators under 'Data Access' and 'Files', with 'Read Excel' selected.
- Parameters:** A panel for configuring the 'Process' operator, showing 'logverbosity' set to 'init' and a 'logfile' field.
- Help:** A panel providing information about the 'Process' operator, including a synopsis and description.

The top menu bar includes 'File', 'Edit', 'Process', 'View', 'Connections', 'Cloud', 'Settings', 'Extensions', and 'Help'. The top toolbar contains icons for 'Run' (a play button, circled in red), 'Save', and 'Undo'. The bottom toolbar shows 'Recommended Operators' with a list of operators and their usage percentages: 'Select Attributes' (29%), 'Set Role' (26%), 'Filter Examples' (22%), and 'Apply Model' (20%).

Vizualizace dat



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Software interface for data visualization, showing a grid of chart types and a repository panel.

The interface includes a menu bar (File, Edit, Process, View, Connections, Cloud, Settings, Extensions, Help) and a toolbar with icons for file operations and execution. The main workspace displays a grid of chart types under the heading "ExampleSet (Read Excel)".

On the left, a sidebar lists categories: Data, Statistics, Charts, Advanced Charts, and Annotations. The "Data" category is selected, and a dropdown menu shows "Pie" highlighted with a red circle. Above the chart grid, a legend indicates data distribution: "vysoké (4)" (blue), "nízké (4)" (green), and "střední (4)" (red).

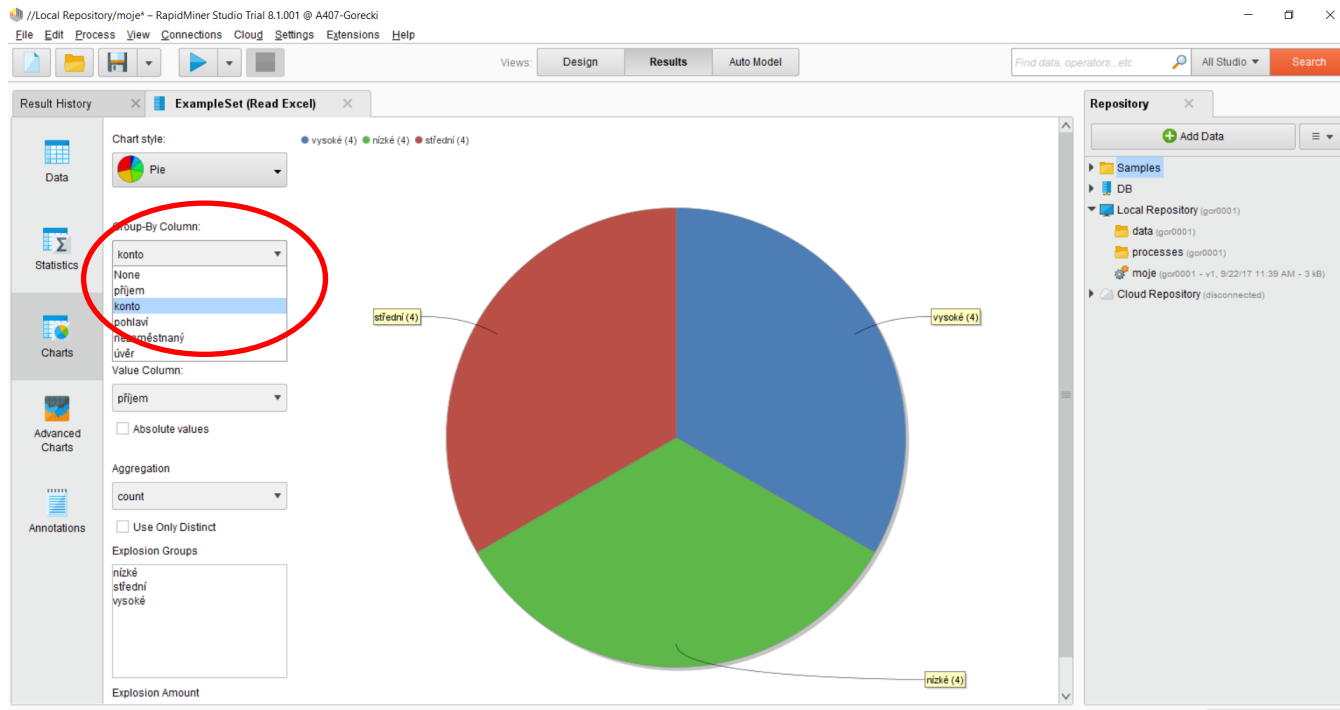
The chart grid contains various visualization options such as Scatter, Series, Histogram, and Pie. The "Pie" chart type is highlighted with a red circle. A yellow label "vysoké (4)" is positioned near the top right of the grid, and another yellow label "nízké (4)" is near the bottom right.

On the right, a "Repository" panel shows a tree view of data sources: Samples, DB, Local Repository (igor0001), and Cloud Repository (disconnected). Under "Local Repository", there are folders for "data" and "processes", and a file named "moje".

Vizualizace dat



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ



Rozhodovací stromy – Volba cílového atributu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

The screenshot displays the RapidMiner Studio interface with the following components:

- Repository:** Shows a tree view of data sources including 'Samples', 'DB', 'Local Repository', and 'Cloud Repository'.
- Process:** A workflow diagram showing a 'Read Excel' operator connected to a 'Set Role' operator. Both operators are circled in red.
- Parameters:** A configuration panel for the 'Set Role' operator. The 'attribute name' dropdown is set to 'úvěr' and the 'target role' dropdown is set to 'label'. Both dropdowns are circled in red.
- Operators:** A search results panel showing 'set ro' as a search term. The 'Set Role' operator is highlighted with a red circle.
- Help:** A help panel for the 'Set Role' operator, providing tags and a synopsis.
- Recommended Operators:** A bottom panel suggesting other operators: Retrieve (70%), Select Attributes (44%), Apply Model (28%), and Filter Examples (26%).

Parameters

Set Role

attribute name: úvěr

target role: label

set additional roles: Edit List (0)...

Change compatibility (8.1.001)

Help

Set Role
RapidMiner Studio Core

Tags: Label, Target, Id, Class, Dependent, Independent, Special, Regular, Inputs, Columns, Attributes, Features, Variables, Types, Names & Roles

Synopsis

This Operator is used to change the role of one or more Attributes.

[Jump to Tutorial Process](#)

Rozhodovací stromy – Nastavení parametrů modelování



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

The screenshot displays the RapidMiner Studio interface. The main workspace shows a workflow with three operators: Read Excel, Set Role, and Decision Tree. The Decision Tree operator is selected, and its parameters are visible in the right-hand panel. The parameters are:

- Decision Tree
- criterion: gain_ratio
- maximal depth: 20
- apply pruning:
- confidence: 0.25
- apply prepruning:
- minimal gain: 0.1
- minimal leaf size: 1
- minimal size for split: 2
- Hide advanced parameters:

The interface also shows the Repository panel on the left, the Operator palette at the bottom left, and the Help panel at the bottom right. The Decision Tree help panel includes the following text:

Decision Tree
Concurrency

Tags: Supervised, Classification, Regression, Model, Trees

Synopsis
This Operator generates a decision tree model, which can be used for classification and regression.

[Jump to Tutorial Process](#)

Rozhodovací stromy – Výsledek



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Local Repository/moje* – RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc

All Studio Search

Result History Tree (Decision Tree)

Graph

Zoom

Tree

Node Labels

Edge Labels

Description

Annotations

```
graph TD; A[přijem] -- vysoký --> B[ano]; A -- nízký --> C[konto]; C -- vysoké --> D[ano]; C -- střední --> E[nezaměstnaný]; C -- nízké --> F[ne]; E -- ano --> G[ano]; E -- ne --> H[ne]; H -- ano --> I[ano]; H -- ne --> J[ne];
```

Repository

Add Data

- Samples
- DB
- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - moje (gor0001 - v1, 9/22/17 11:39 AM - 3 KB)
- Cloud Repository (disconnected)

ne

Distribution: 0 ano, 2 ne

Number of items: 2

Ratio of total: 16.67%

Rozhodovací stromy – Výsledek (pravidla)



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Result History x Tree (Decision Tree) x

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Graph

Description

Annotations

```
příjem = nízký
| konto = nízké: ne (ano=0, ne=2)
| konto = střední
| nezaměstnaný = ano: ne (ano=0, ne=2)
| nezaměstnaný = ne: ano (ano=1, ne=0)
| konto = vysoké: ano (ano=2, ne=0)
příjem = vysoký: ano (ano=3, ne=0)
```

Repository

Add Data

- Samples
- DB
- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - moje (gor0001 - v1, 9/22/17 11:39 AM - 3 kB)
- Cloud Repository (disconnected)

Rozhodovací stromy – Úprava parametrů modelu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Local Repository/moje* – RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connection Cloud Settings Extensions Help

Views: Design Results Auto Model Find data, operators...etc All Studio Search

Repository

- DB
- Local Repository (gor001)
 - data (gor001)
 - processes (gor001)
 - moje (gor001 - v1, 9/22/17 11:39 AM - 3 kB)
- Cloud Repository (disconnected)

Process

Process

100%

Read Excel → Set Role → Decision Tree

Parameters

Decision Tree

- criterion: gain_ratio
- maximal depth: 20
- apply pruning:
- confidence: 0.25
- apply prepruning:
- minimal gain: 0.1
- minimal leaf size: 1
- minimal size for split: 4

[Hide advanced parameters](#)

Help

Decision Tree

Concurrency

Tags: Supervised, Classification, Regression, Model, Trees

Synopsis

This Operator generates a decision tree model, which can be used for classification and regression.

[Jump to Tutorial Process](#)

Operators

deci

- Predictive (8)
 - Trees (8)
 - Decision Tree
 - Random Forest
 - Gradient Boosted Trees
 - ID3
 - Decision Stump
 - Decision Tree (Multiway)
 - Decision Tree (Weight-Based)

We found "Weka Extension" and "WhiBo" in the Marketplace [Show me!](#)

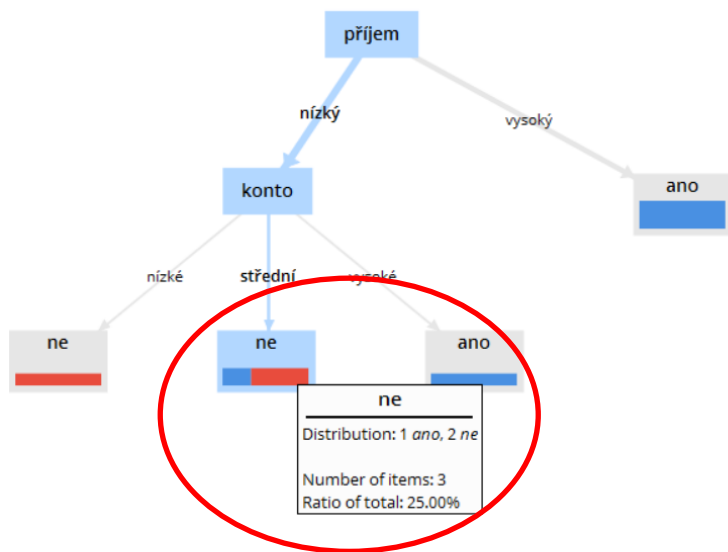
Recommended Operators

- Retrieve 73%
- Select Attributes 39%
- Apply Model 36%
- Filter Examples 27%

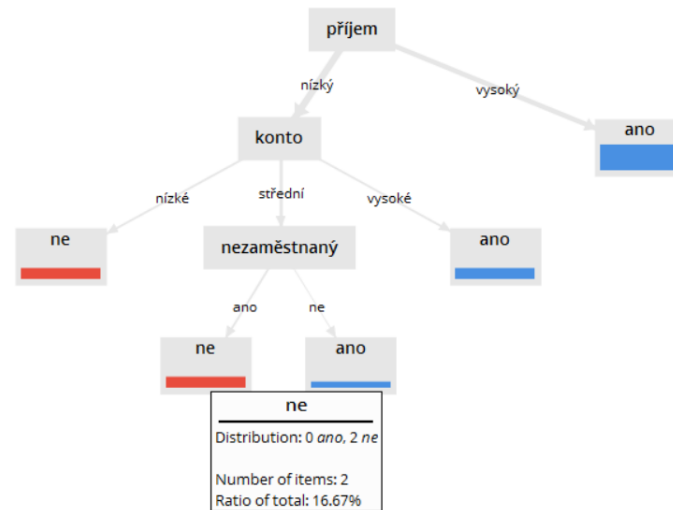
Rozhodovací stromy – Výsledek 2



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Výsledek 2



Výsledek 1

Banka - Testování modelu – Křížová validace



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

://Local Repository/processes/stromy_bank_a_crossval* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators, etc. All Studio Search

Repository

- Samples
- DB
- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - stromy (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_crossval (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
 - moje (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
- Cloud Repository (disconnected)

Process

Process

100%

Process flow: **Read CSV** → **Set Role** → **Cross Validation**

Parameters

Read CSV

- Import Configuration Wizard
- csv file: data/Bank/bank.csv
- column separators: ;
- trim lines:
- use quotes:
- quotes character: "
- escape character: \
- Hide advanced parameters
- Change compatibility (8.1.001)

Help

Read CSV

RapidMiner Studio Core

Tags: Load, Import, Read, Data, Files, Text, Commas, Spreadsheet, Excel, Datasets, Tsv

Synopsis

This Operator reads an ExampleSet from the specified CSV file.

[Jump to Tutorial Process](#)

Operators

cross va

- Validation (1)
 - Cross Validation

No results were found.

Recommended Operators

- Retrieve 68%
- Select Attributes 42%
- Apply Model 27%
- Filter Examples 23%

Banka - Testování modelu – Křížová validace



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

///Local Repository/processes/stromy_bank_a_crossval* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model Find data, operators, etc. All Studio Search

Repository

- Local Repository (igor0001)
 - data (igor0001)
 - processes (igor0001)
 - stromy (igor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_crossval (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)
 - moje (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)

Operators

cross va

- Validation (1)
 - Cross Validation

Process

Process: Cross Validation

100%

Training: Decision Tree

Testing: Apply Model, Performance

Parameters

Decision Tree

- criterion: information_gain
- maximal depth: 20
- apply pruning:
- confidence: 0.25
- apply prepruning:
- minimal gain: 0.1
- minimal leaf size: 10
- minimal size for split: 20

Help

Decision Tree

Concurrency

Tags: Supervised, Classification, Regression, Model, Trees

Synopsis

This Operator generates a decision tree model, which can be used for classification and regression.

[Jump to Tutorial Process](#)

Recommended Operators

- Performance (CL... 33%
- Performance (Bl... 22%
- k-NN 17%
- Naive Bayes 15%

Banka - Testování modelu – Křížová validace - výsledky



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Local Repository/processes/stromy_bank_a_crossval* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators, etc. All Studio Search

Result History PerformanceVector (Performance)

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

Table View Plot View

accuracy: 88.43% +/- 0.13% (mikro: 88.43%)

	true no	true yes	class precision
pred. no	3996	519	88.50%
pred. yes	4	2	33.33%
class recall	99.90%	0.38%	

Téměř nepokrývá třídu yes

Repository: Add Data, Samples, DB, Local Repository (igor001), data (igor001), processes (igor001), stromy (igor001 - v1, 4/17/18 3:11 PM - 5 kB), stromy_bank_a_crossval (igor001 - v1, 4/17/18 3:10 PM - 5 kB), moje (igor001 - v1, 4/17/18 3:10 PM - 5 kB), Cloud Repository (disconnected)

Banka - Testování modelu – Křížová validace 2



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

//Local Repository/processes/stromy_bank_a_crossval* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators... All Studio Search

Repository

- DB
- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - stromy (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_crossval (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
 - moje (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
- Cloud Repository (disconnected)

Process

Process: Cross Validation

100%

Training: Decision Tree

Testing: Apply Model, Performance

Parameters

Decision Tree

- criterion: gini_index
- maximal depth: 20
- apply pruning:
- confidence: 0.25
- apply prepruning:
- minimal gain: 0.1
- minimal leaf size: 10
- minimal size for split: 20

[Hide advanced parameters](#)

Help

Decision Tree

Concurrency

Tags: Supervised, Classification, Regression, Model, Trees

Synopsis

This Operator generates a decision tree model, which can be used for classification and regression.

[Jump to Tutorial Process](#)

Recommended Operators

- Performance (CL... 33%
- Performance (BI... 22%
- k-NN 17%
- Naive Bayes 15%

Banka - Testování modelu – Křížová validace 2



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

//Local Repository/processes/stromy_banka_crossval* – RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Result History PerformanceVector (Performance)

Criterion
accuracy
precision
recall
AUC (optimistic)
AUC
AUC (pessimistic)

Description

Annotations

Table View Plot View

accuracy: 88.85% +/- 0.74% (mikro: 88.85%)

	true no	true yes	class precision
pred. no	3869	373	91.21%
pred. yes	131	148	53.05%
class recall	96.73%	28.41%	

Repository

Add Data

- Samples
- DB
- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - stromy (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banka_crossval (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - moje (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
- Cloud Repository (disconnected)

Mnohem lepší model
(parametrizace)
vzhledem k pokrytí
třídy *yes*

Optimalizace parametrů



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

///Local Repository/processes/stromy_bank_a_optimization* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc

All Studio Search

Repository

+ Add Data

- Samples
- DB
- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - stromy (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_crossval (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_optimization (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - moje (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
- Cloud Repository (disconnected)

Process

Process

100%

Process

inp

Read CSV

Set Role

Optimize Parameters (Grid)

res

res

res

Parameters

Optimize Parameters (Grid)

Edit Parameter Settings...

error handling: fail on error

log performance

log all criteria

synchronize

enable parallel execution

Hide advanced parameters

Help

Optimize Parameters (Grid)

Concurrency

Tags: Iterate, Settings, Grid, Search, Tune, Optimal, Parameters

Synopsis

This Operator finds the optimal values of the selected parameters for the Operators in its subprocess.

[Jump to Tutorial Process](#)

Operators

perform

- Cluster Density Performance
- Item Distribution Performance
- Performance
- Extract Performance
- Combine Performances
- Performance (User-Based)
- Performance (Min-Max)
- Performance In Data

We found "Time Series", "Model Visualization Extension" and 2 more results in the Marketplace. [Show me!](#)

Recommended Operators

- Retrieve 70%
- Select Attributes 41%
- Subprocess 32%
- Apply Model 31%

Optimalizace parametrů



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Local Repository/processes/stromy_banka_optimization* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Repository

- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - stromy (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banka_crossval (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banka_optimization (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)

Process

Process > Optimize Parameters (Grid)

Optimize Parameters (Grid)

100%

inp

inp

Cross Validation

mod

per

mod

out

Parameters

Optimize Parameters (Grid)

Edit Parameter Settings...

error handling: fail on error

log performance

log all criteria

synchronize

enable parallel execution

Hide advanced parameters

Help

Optimize Parameters (Grid)

Concurrency

Tags: Iterate, Settings, Grid, Search, Tune, Optimal, Parameters

Synopsis

This Operator finds the optimal values of the selected parameters for the Operators in its subprocess.

[Lump to Tutorial Process](#)

Operators

perform

- Cluster Density Performance
- Item Distribution Performance
- Performance
- Extract Performance
- Combine Performances
- Performance (User-Based)
- Performance (Min-Max)
- Performance to Data

We found "Time Series", "Model Visualization Extension" and 2 more results in the Marketplace. [Show me!](#)

Recommended Operators

- Log 61%
- Apply Model 5%
- Sample 5%
- Multiply 4%

Optimalizace parametrů



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

://Local Repository/processes/stromy_banko_optimization* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Repository

- Local Repository (igor0001)
 - data (igor0001)
 - processes (igor0001)
 - stromy_igor0001 - v1, 4/17/18 3:11 PM - 5 kB
 - stromy_banko_crossval (igor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banko_optimization (igor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - moje (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)

Operators

perfor

- Cluster Density Performance
- Item Distribution Performance
- Performance
- Extract Performance
- Combine Performances
- Performance (User-Based)
- Performance (Min-Max)
- Performance to Data

We found "Time Series", "Model Visualization Extension" and 2 more results in the Marketplace. [Show me!](#)

Process: Optimize Parameters (Grid) > Cross Validation

100%

Decision Tree

Apply Model

Performance

Parameters

Cross Validation

- split on batch attribute
- leave one out
- number of folds: 10
- sampling type: automatic
- use local random seed
- enable parallel execution

[Hide advanced parameters](#)

Help

Cross Validation

Concurrency

Tags: Cross-Validations, Cross-validations, Folds, K-Folds, K-folds, Validations, Estimations, Evaluations, Performances, Splitting, X-Validation, X-Prediction, Validation

Synopsis

This Operator performs a cross validation to estimate the statistical performance of a learning model.

[Jump to Tutorial Process](#)

Recommended Operators

- Performance (CL... 33%
- Performance (Bl... 22%
- k-NN 17%
- Naive Bayes 15%

Optimalizace parametrů



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Local Repository/processes/stromy_banka_optimization* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Repository

- Local Repository (igor0001)
 - data (igor0001)
 - processes (igor0001)
 - stromy (igor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banka_crossval (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)
 - stromy_banka_optimization (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)
 - moje (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)

Operators

performer

- Cluster Density Performance
- Item Distribution Performance
- Performance
- Extract Performance
- Combine Performances
- Performance (User-Based)
- Performance (Min-Max)
- Performance to Data

We found "Time Series", "Model Visualization Extension" and 2 more results in the Marketplace. [Show me!](#)

Process

Process

100%

Process

inp

Read CSV

Set Role

Optimize Parameters (Grid)

per

res

res

res

res

res

Parameters

Optimize Parameters (Grid)

Edit Parameter Settings...

error handling fail on error

log performance

log all criteria

synchronize

enable parallel execution

Hide advanced parameters

Help

Optimize Parameters (Grid)

Concurrency

Tags: Iterate, Settings, Grid, Search, Tune, Optimal, Parameters

Synopsis

This Operator finds the optimal values of the selected parameters for the Operators in its subprocess.

[Jump to Tutorial Process](#)

Select Parameters: configure operator

Select Parameters: configure operator

Configure this operator by means of a Wizard.

Operators

- Cross Validation (Cross Validation)
- Decision Tree (Decision Tree)
- Apply Model (Apply Model)
- Performance (Performance)

Parameters

Selected Parameters

- Decision Tree.minimal_leaf_size
- Decision Tree.criterion

Grid/Range

Min	Max	Steps	Scale
10	500	50	linear

Value List

- gain_ratio
- information_gain
- gini_index
- accuracy

Recent

- least_square

2 parameters / 204 combinations selected

Grid List

OK Cancel

Optimalizace parametrů



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

///Local Repository/processes/stromy_banka_optimization* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Repository

- Local Repository (igor0001)
 - data (igor0001)
 - processes (igor0001)
 - stromy (igor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banka_crossval (igor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_banka_optimization (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)
 - moje (igor0001 - v1, 4/17/18 3:10 PM - 5 kB)

Process

Process

100%

Process

inp

Read CSV

Set Role

Optimize Parameters (Grid)

res

res

res

res

res

Select Parameters: configure operator

Select Parameters: **configure operator**
Configure this operator by means of a Wizard.

Operators

- Cross Validation (Cross Validation)
- Decision Tree (Decision Tree)
- Apply Model (Apply Model)
- Performance (Performance)

Parameters

Selected Parameters

- Decision Tree_minimal_leaf_size
- Decision Tree_criterion

Grid/Range

Min	Max	Steps	Scale
10	500	50	linear

-Value List

- 10
- 20
- 30
- 39

Recommended

Grid List 2 parameters / 204 combinations selected

OK Cancel

Parameters

Optimize Parameters (Grid)

Edit Parameter Settings...

error handling fail on error

log performance

log all criteria

synchronize

enable parallel execution

Hide advanced parameters

Help

Optimize Parameters (Grid)

Concurrency

Tags: Iterate, Settings, Grid, Search, Tune, Optimal, Parameters

Synopsis

This Operator finds the optimal values of the selected parameters for the Operators in its subprocess.

[Jump to Tutorial Process](#)

Optimalizace parametrů - výsledek



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

//Local Repository/processes/stromy_bank_a_optimization* - RapidMiner Studio Trial 8.1.001 @ A407-Gorecki

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Auto Model

Find data, operators...etc All Studio Search

Tree (Decision Tree) PerformanceVector (Performance) Optimize Parameters (Grid)

Result History ParameterSet (Optimize Parameters (Grid))

ParameterSet

Description

Parameter set:

Performance:

PerformanceVector [

----accuracy: 89.01% +/- 0.67% (mikro: 89.01%)

ConfusionMatrix:

True:	no	yes
no:	3855	352
yes:	145	169

----precision: 59.80% +/- 12.06% (mikro: 53.82%) (positive class: yes)

ConfusionMatrix:

True:	no	yes
no:	3855	352
yes:	145	169

----recall: 32.40% +/- 17.80% (mikro: 32.44%) (positive class: yes)

ConfusionMatrix:

True:	no	yes
no:	3855	352
yes:	145	169

----AUC (optimistic): 0.908 +/- 0.016 (mikro: 0.908) (positive class: yes)

----AUC: 0.747 +/- 0.041 (mikro: 0.747) (positive class: yes)

----AUC (pessimistic): 0.586 +/- 0.067 (mikro: 0.586) (positive class: yes)

]

Decision Tree.minimal_leaf_size = 49

Decision Tree.criterion = gini_index

Annotations

Repository

- Local Repository (gor0001)
 - data (gor0001)
 - processes (gor0001)
 - stromy (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_crossval (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - stromy_bank_a_optimization (gor0001 - v1, 4/17/18 3:11 PM - 5 kB)
 - moje (gor0001 - v1, 4/17/18 3:10 PM - 5 kB)
- Cloud Repository (disconnected)

Děkuji za pozornost