

GUIDED ANALYTICS LEARNATHON



@ODSC
@KNIME

Boston | April 30 - May 4, 2019



BOSTON

APR 30 – MAY 3

Guided Analytics Learnathon: Building Applications for Automated Machine Learning

Paolo Tamagnini

Data Scientist,
KNIME, Inc.





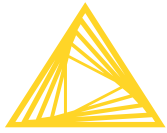
BOSTON
APR 30 – MAY 3

Guided Analytics Learnathon: Building Applications for Automated Machine Learning

Scott Fincher

Data Scientist,
KNIME, Inc.





Open for Innovation[®]

KNIME

Guided Analytics Learnathon

Sessions

Introduction to KNIME Analytics Platform and KNIME Server

Building Applications for Automated Machine Learning

Let's build a guided analytics workflow!

Wrap-up Session and Demo of Solution



#KNIME
#ODSCEast

Paolo Tamagnini, KNIME
Scott Fincher, KNIME



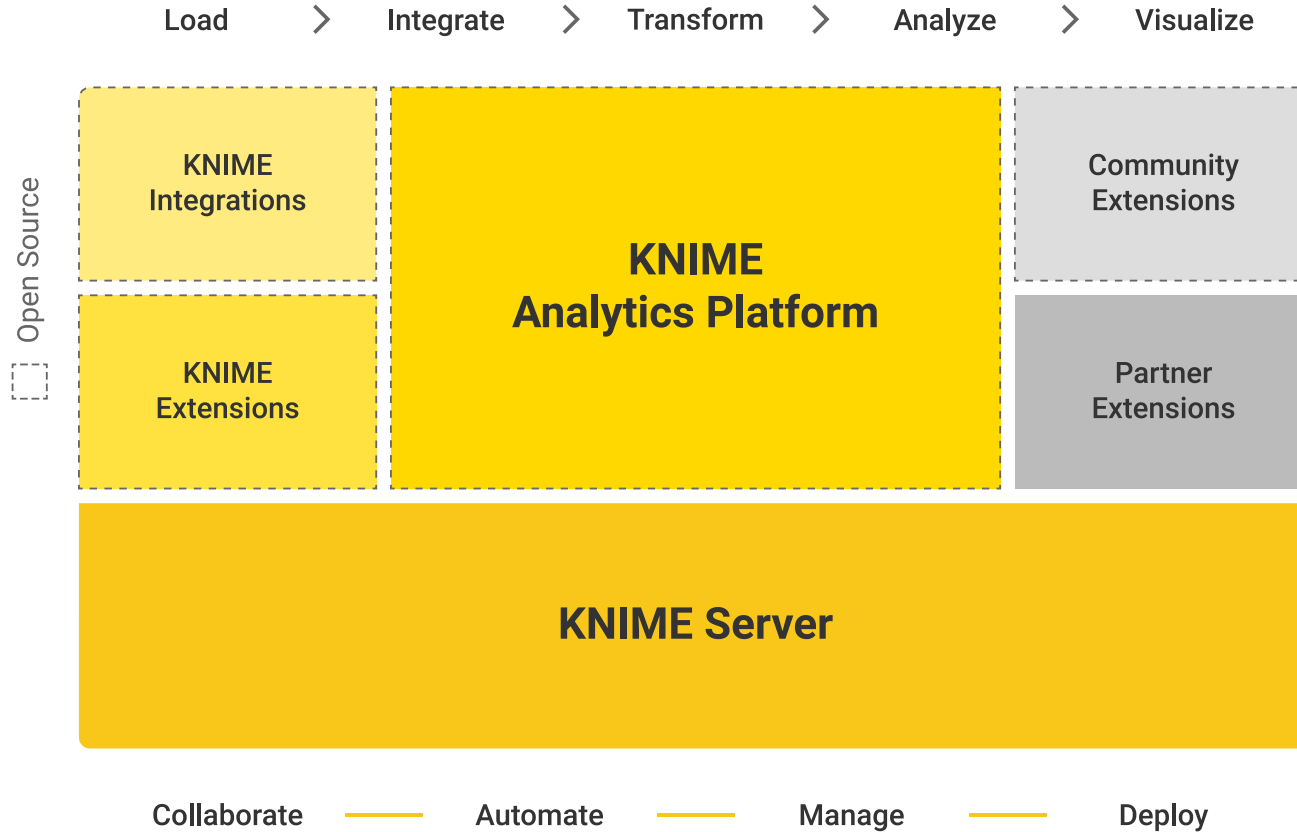
Introduction to KNIME Analytics Platform and KNIME Server

Paolo Tamagnini, KNIME



#KNIME
#ODSCEast

KNIME® Software





Guided Analytics Learnathon

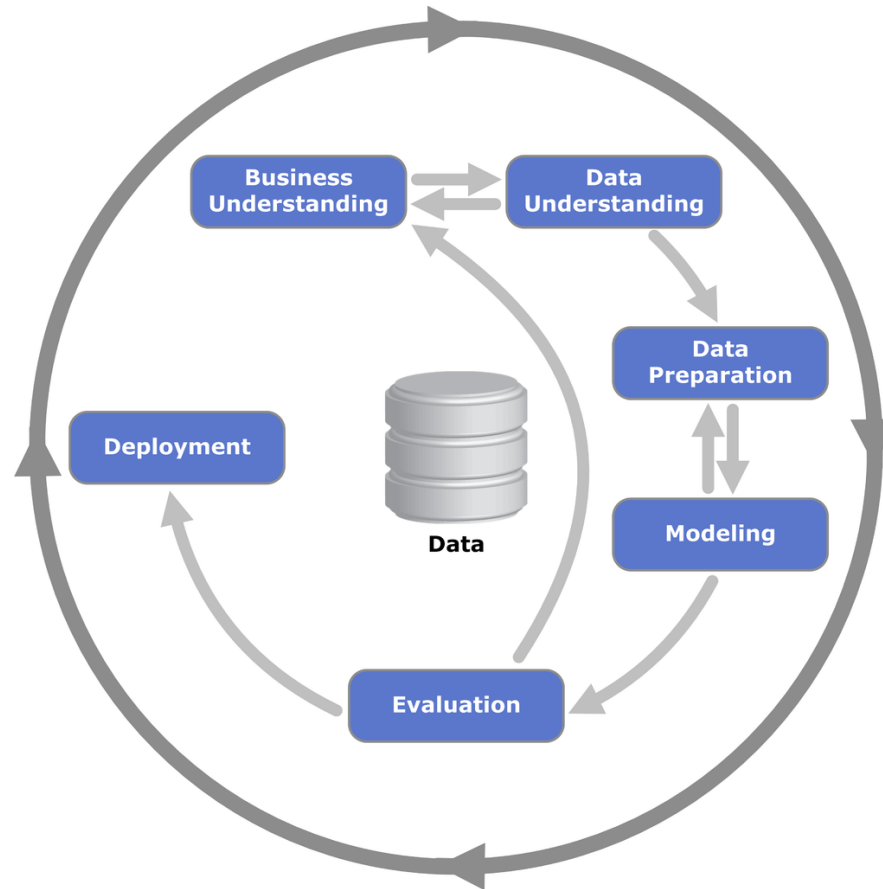
Building Applications for Automated ML

Paolo Tamagnini, KNIME



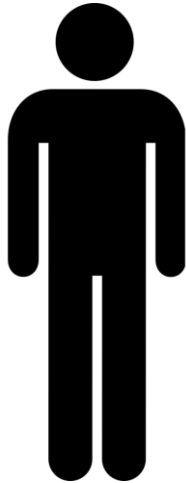
#KNIME
#ODSCEast

The CRISP-DM Cycle

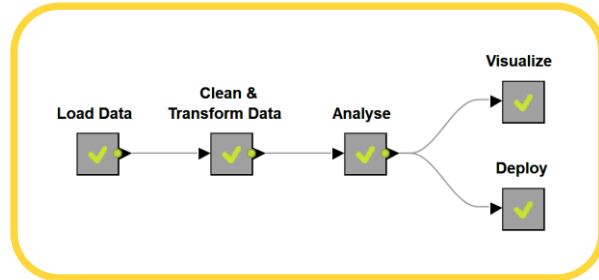


In theory..

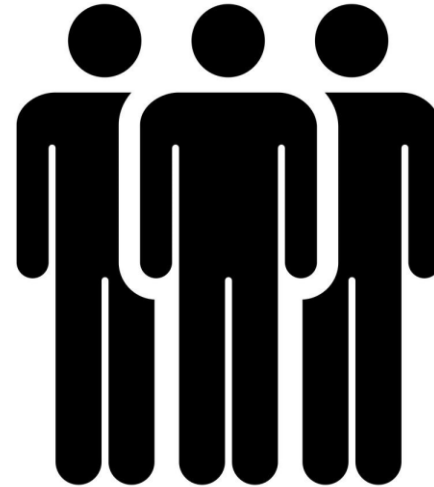
tadaahh!



Data Scientist



*Great! Good Job!
Awesome!*



Business Analysts

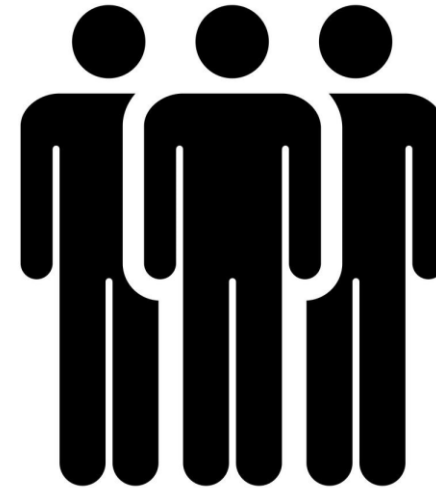
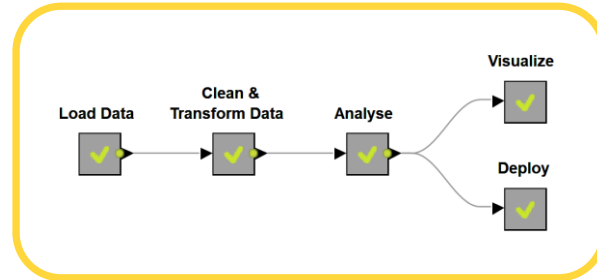
In practice..

*Can I use a **neural network** instead of a **decision tree**?*
*Can I **upload** my own file? Can I optimize the **model parameters**?*
*Would it be possible to add some **feature engineering**?*

!?

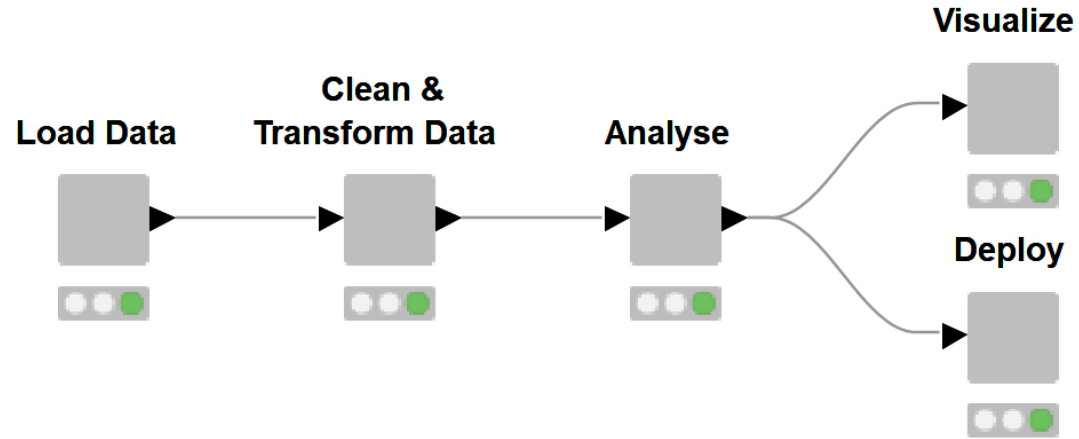


Data Scientist

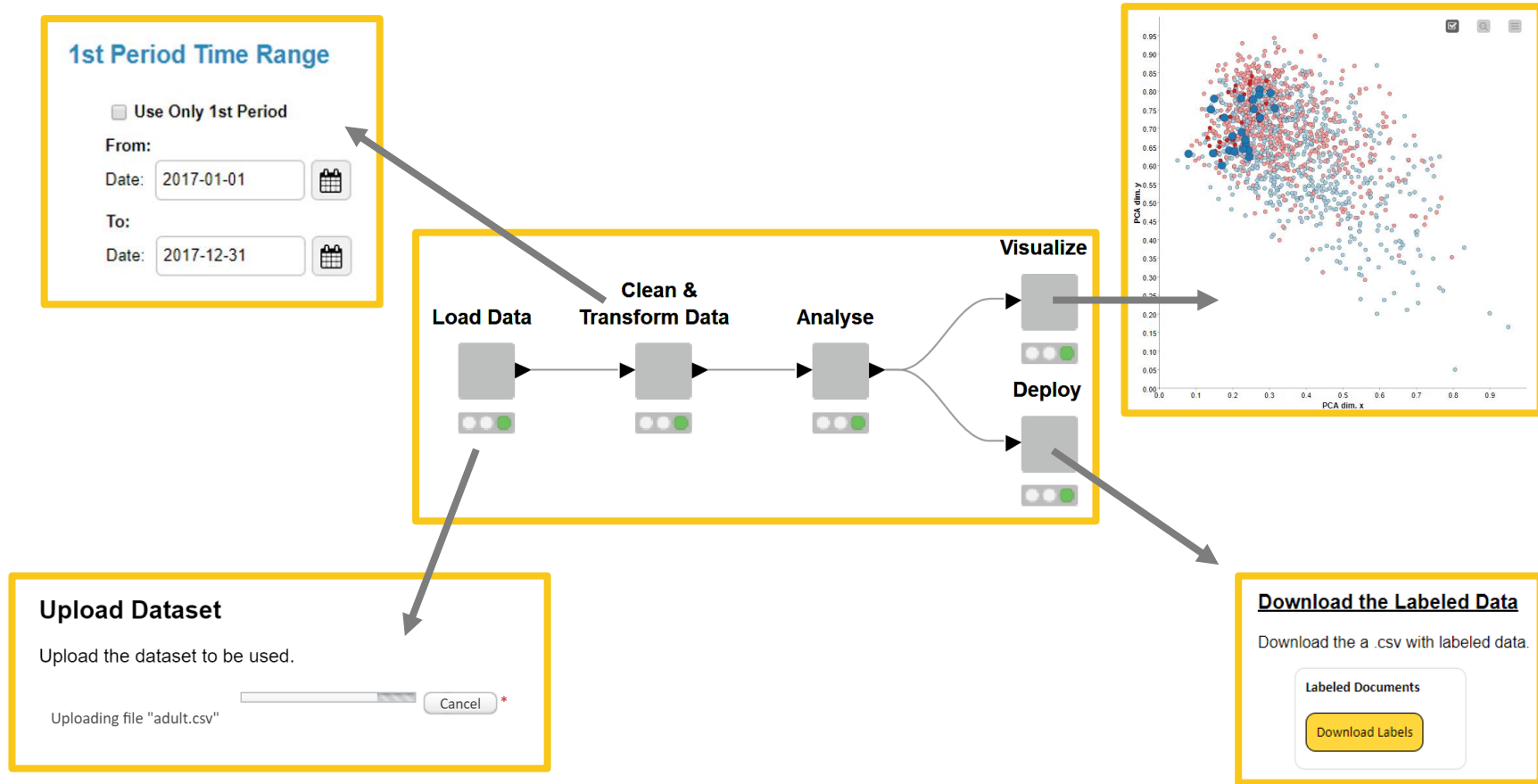


Business Analysts

The CRISP-DM Cycle with Wrapped Metanodes



Examples of Wrapped Metanode Views

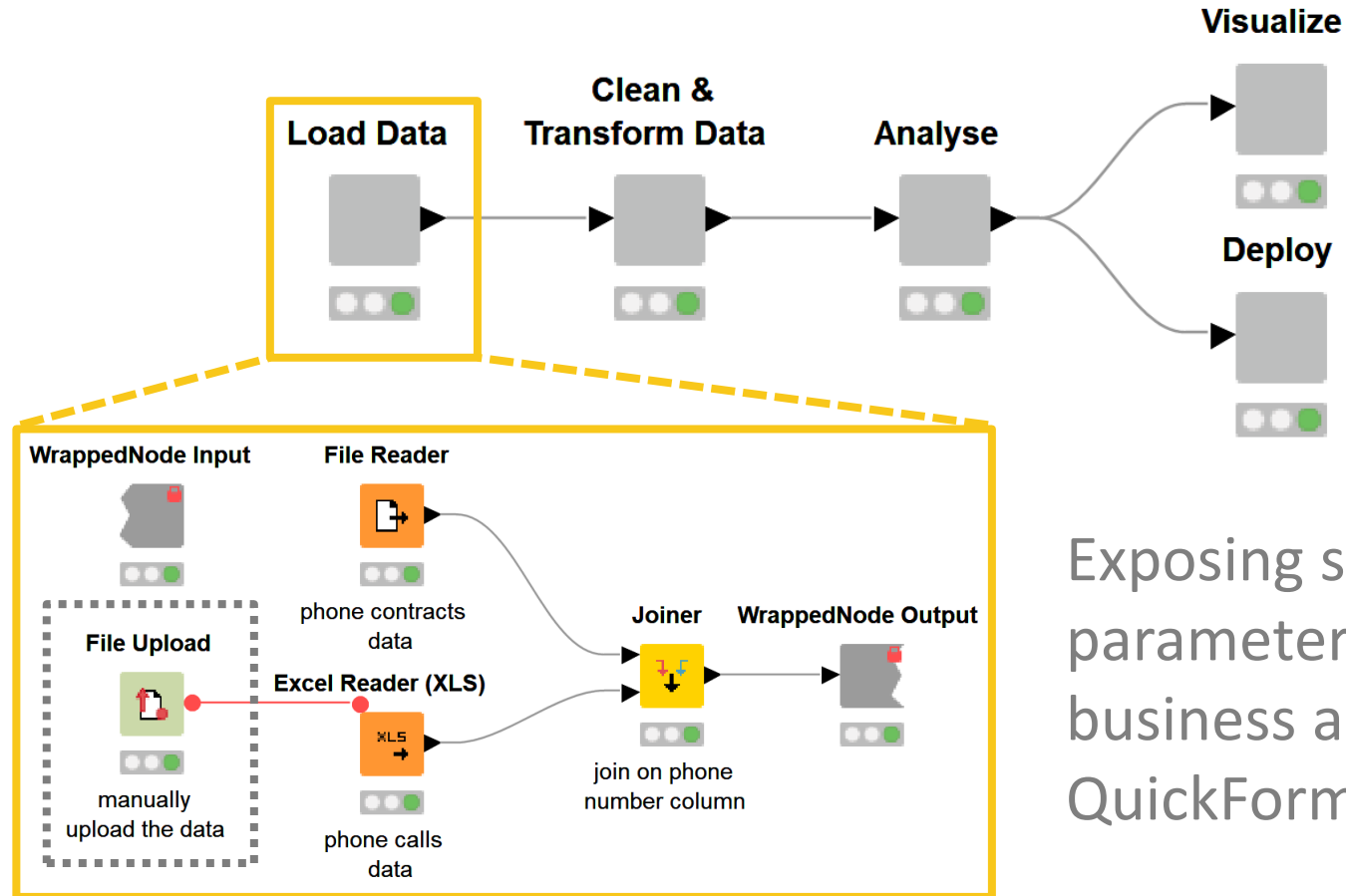


Quickforms and JavaScript Nodes

- Workflow Control
 - Quickforms
 - Input
 - Boolean Input
 - Credentials Input
 - Date&Time Input
 - Double Input
 - File Chooser
 - File Upload
 - Integer Input
 - List Box Input
 - Molecule String Input
 - Slider Input
 - String Input
 - JS Autocomplete Text Input
 - Selection
 - Column Selection
 - Multiple Selections
 - Single Selection
 - Value Selection
 - Filter
 - Column Filter
 - Range Slider Filter Definition
 - Value Filter

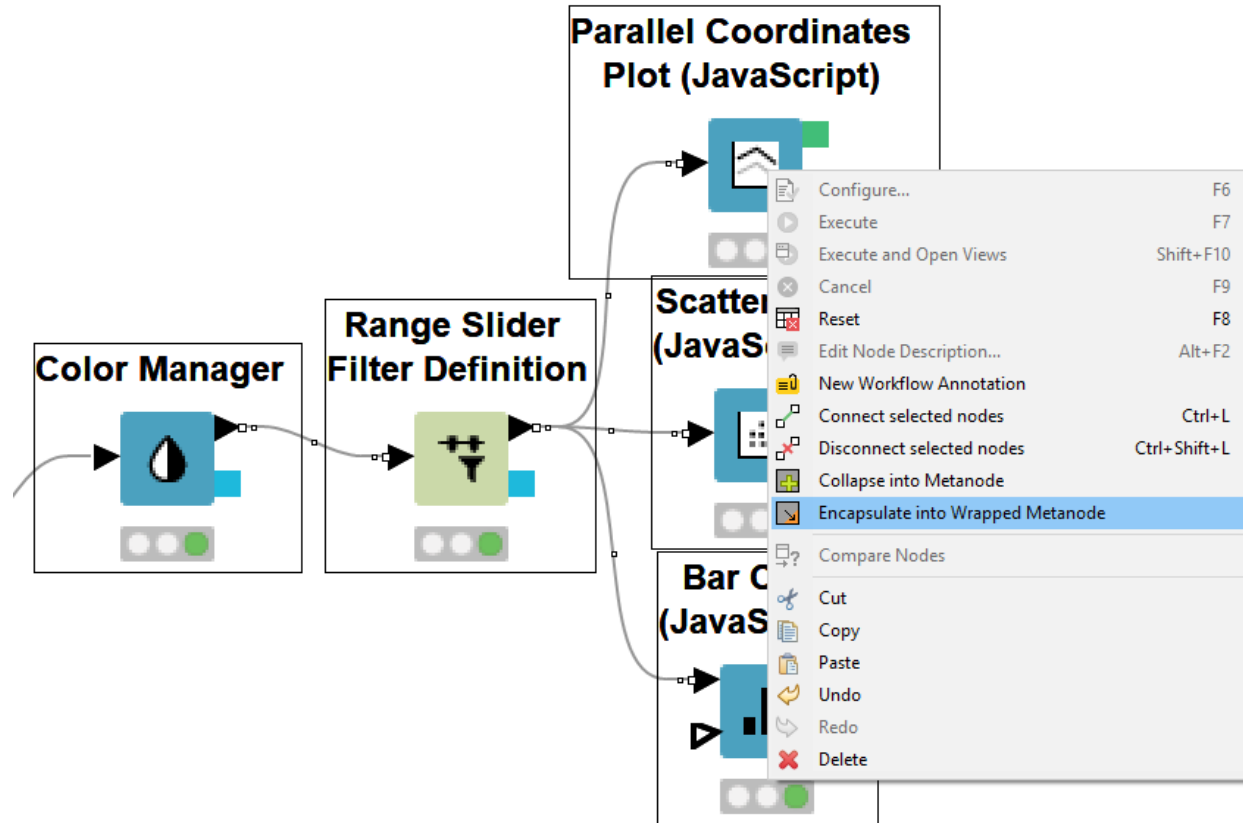
- Views
 - JavaScript
 - Network Viewer (JavaScript)
 - Table View (JavaScript)
 - Scatter Plot (JavaScript)
 - Line Plot (JavaScript)
 - Lift Chart (JavaScript)
 - ROC Curve (JavaScript)
 - Decision Tree View (JavaScript)
 - JS Generic JavaScript View
 - Bar Chart (JavaScript)
 - Box Plot (JavaScript)
 - Conditional Box Plot (JavaScript)
 - Parallel Coordinates Plot (JavaScript)
 - Pie/Donut Chart (JavaScript)
 - Stacked Area Chart (JavaScript)
 - Sunburst Chart (JavaScript)
 - KNIME Labs
 - JavaScript Views (Labs)
 - Data Explorer (JavaScript)
 - Table Editor (JavaScript)
 - Tag Cloud (JavaScript)

Data Access



Exposing some parameters to the business analyst using QuickForms nodes!

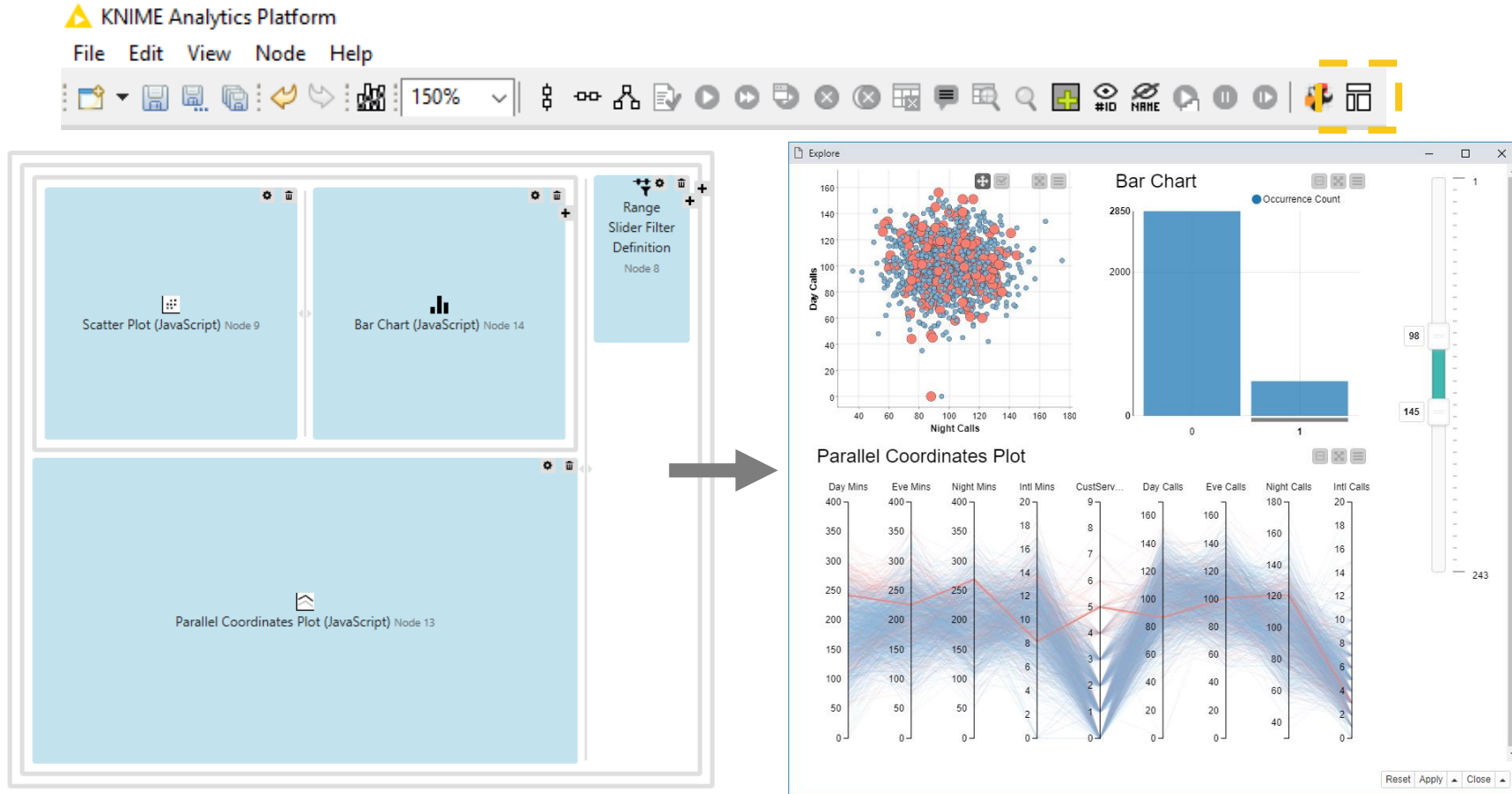
How to Make a Wrapped Metanode



Data Visualization of Results



Customizing the View Layout



Connecting to K

Select New Content

Mounting a new resource for display in the KNIME Explorer

Please select the type of resource that should be mounted.

⌚ Checking Server connection...

- KNIME ServerSpace
- KNIME Community Server

Server name or address:

Server address: **https://learnathon.knime.com**

Use REST: ☒

Server login credentials:

Username: **user-0** Password (optional): **•••••**

Test Connection

Enter the name of the KNIME Server (Mount ID).

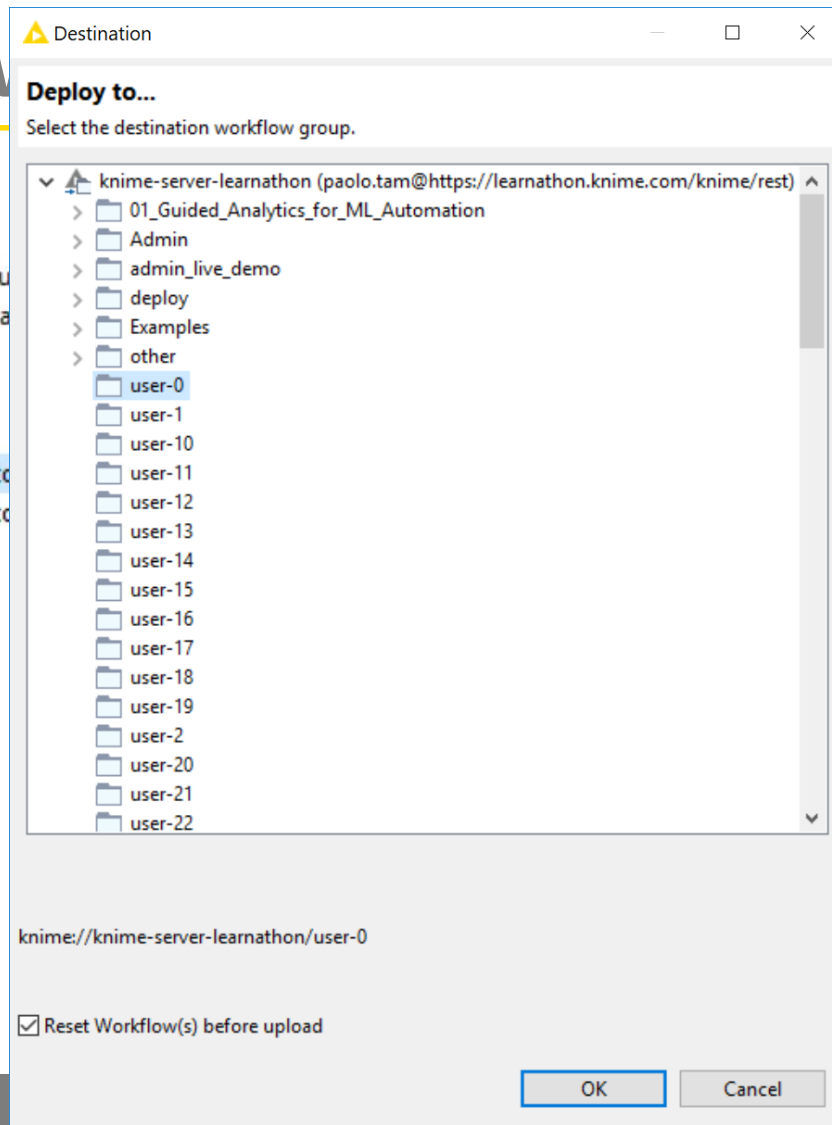
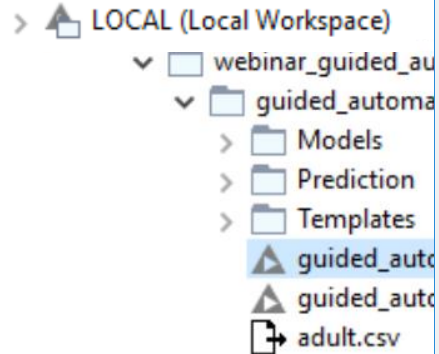
A default is fetched from the server when "Test Connection" is pressed or 'Use REST' is selected.

Mount ID: **knime-server**

Reset Mount ID

OK Cancel

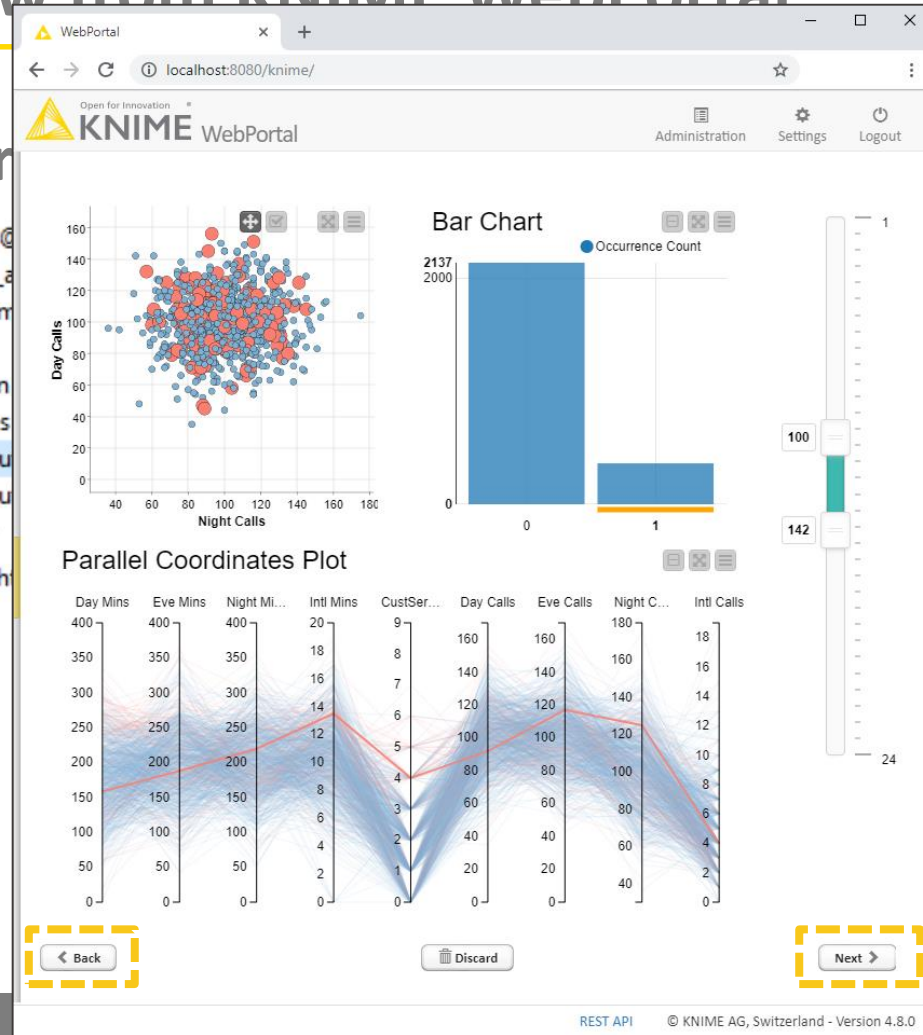
Deploying your V



Composite View from KNIME WebPortal

From the workspace

- > knime-server (knimeadmin@h...)
 - ▼ webinar_guided_a...
 - ▼ guided_autom...
 - > Models
 - > Prediction
 - > Templates
 - ▲ guided_au...
 - ▲ guided_au...
 - ▶ adult.csv
 - > EXAMPLES (knime-guest@h...)
 - > LOCAL (Local Workspace)



Interactivity across Charts: Selection and Filter Events



A Solution to Automated Machine Learning

What is Guided Analytics?

Developing interactive systems assisting the business analyst in finding new insights and predict future outcomes

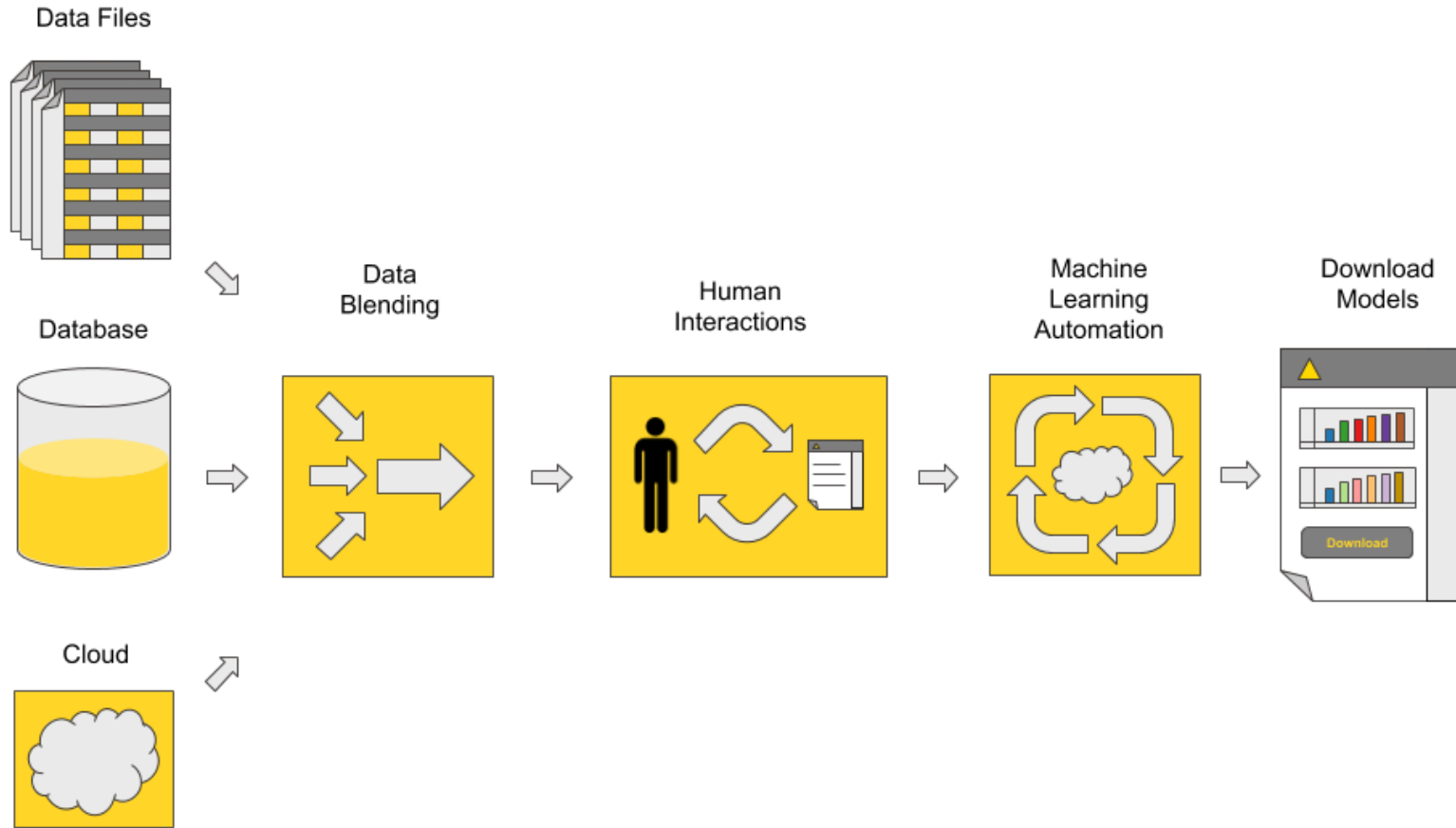
Guided Automation is a special instance of **Guided Analytics** aimed at *intelligently* automating the ML cycle

Two Key Questions

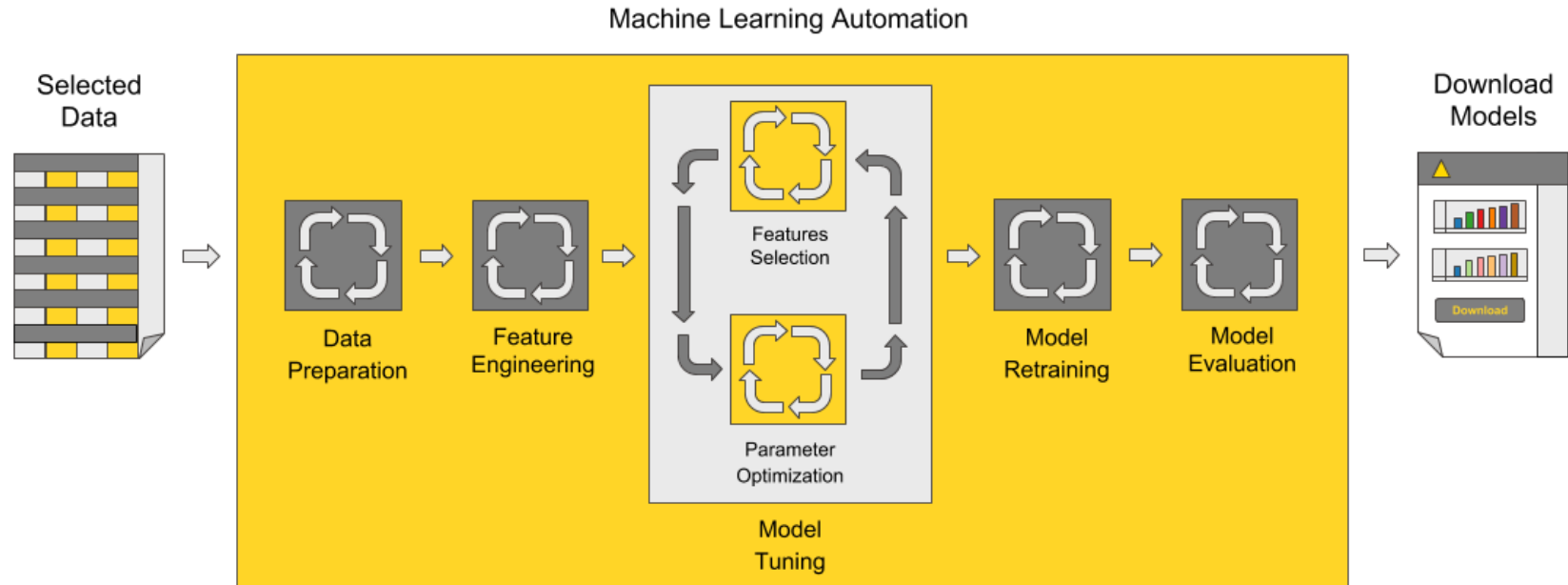
Can we automate the ML cycle in KNIME?

Is it useful to automate the ML cycle?

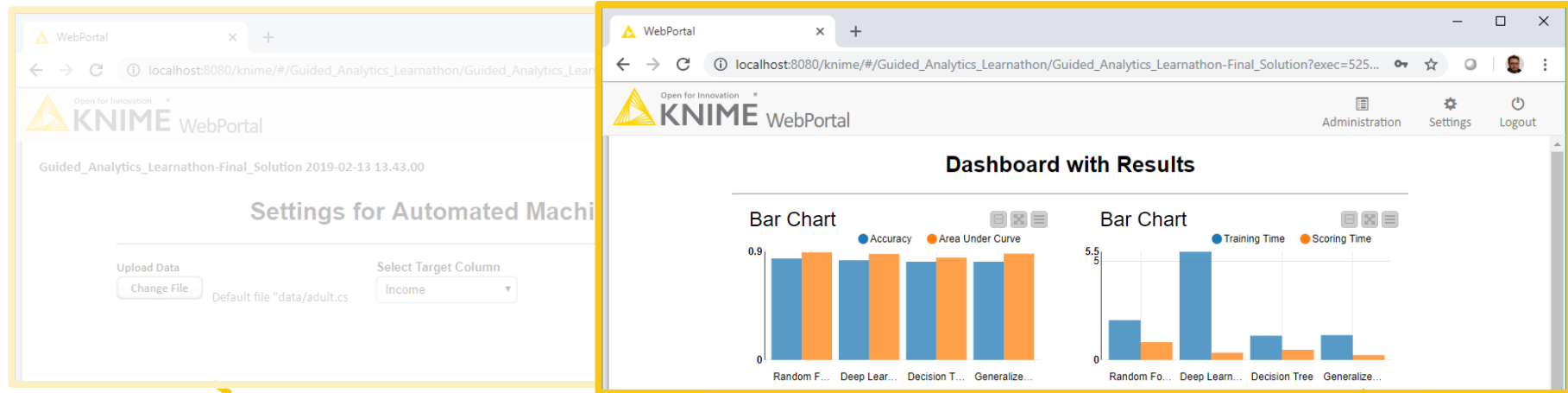
A General Recipe for Guided Automation



Machine Learning Automation



Today Exercise: a simple example of ML automation



Group 1 : Start Web Page

Guided Automation Metanode

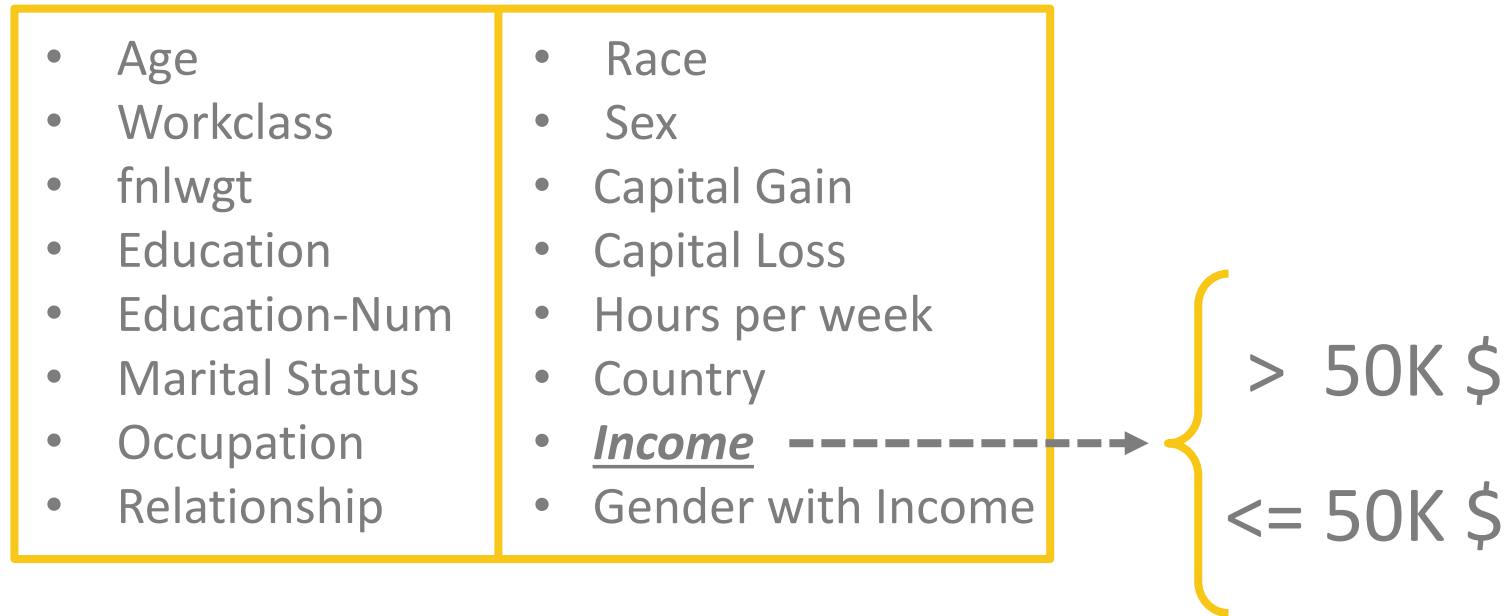
Group 2 : Final Web Page

Ctrl + Double Left Click to Open the Wrapped Metanode

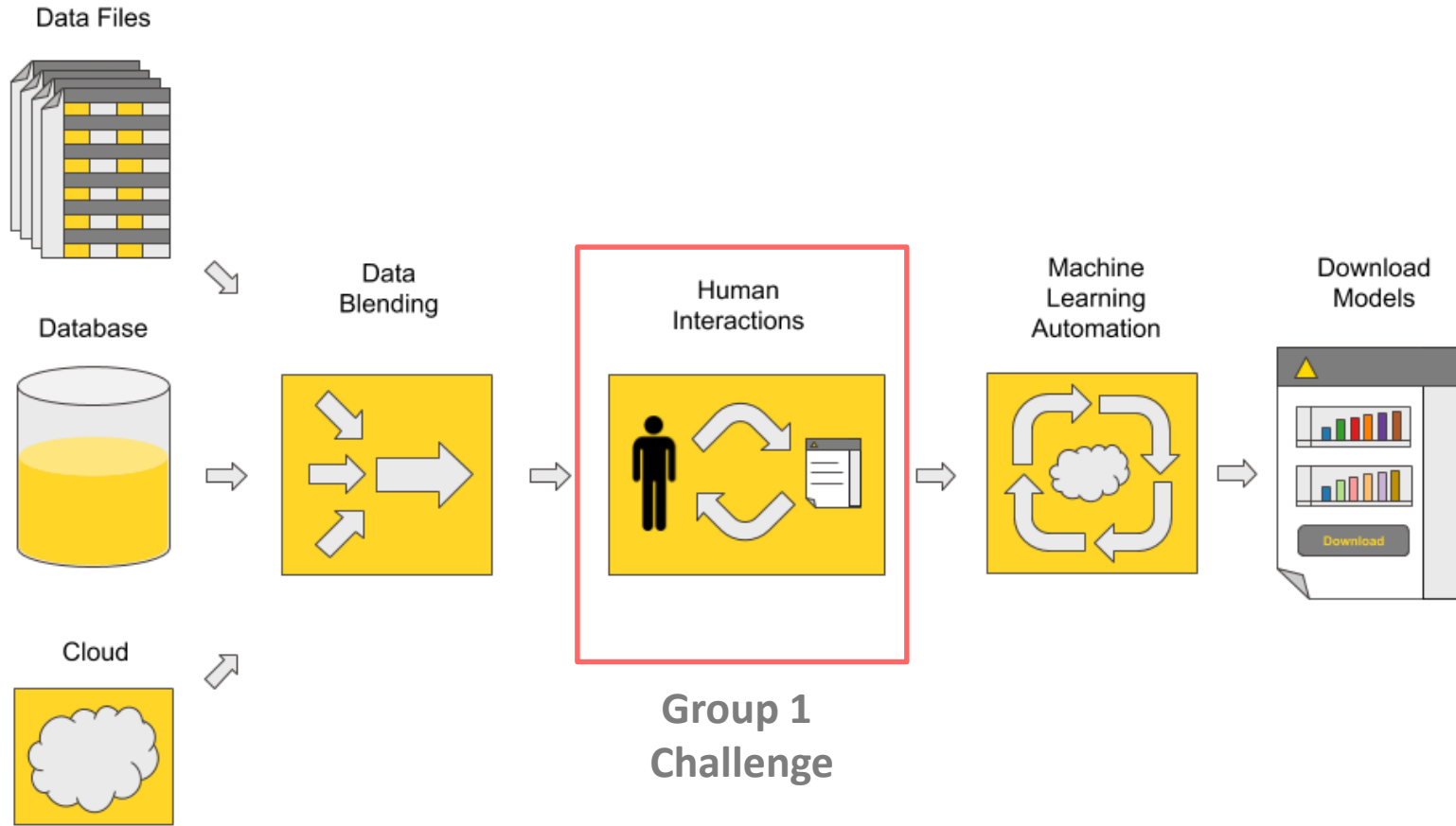
Ctrl + Double Left Click to Open the Wrapped Metanode

Today Exercise: Adult Dataset

Source: archive.ics.uci.edu/ml/datasets/adult

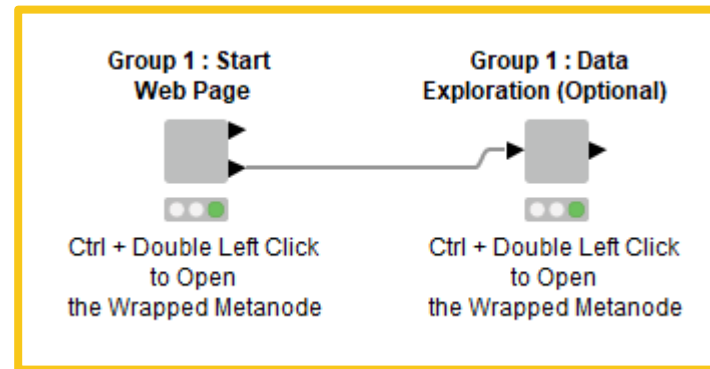
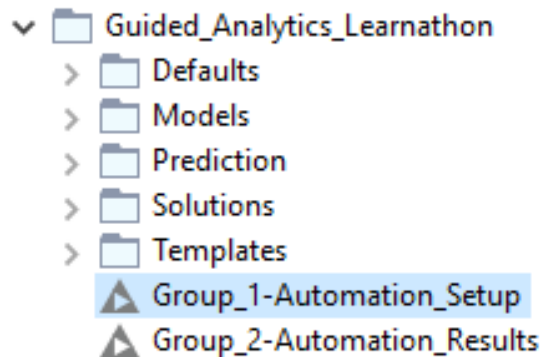


Group 1: User Interface for Automated ML Settings

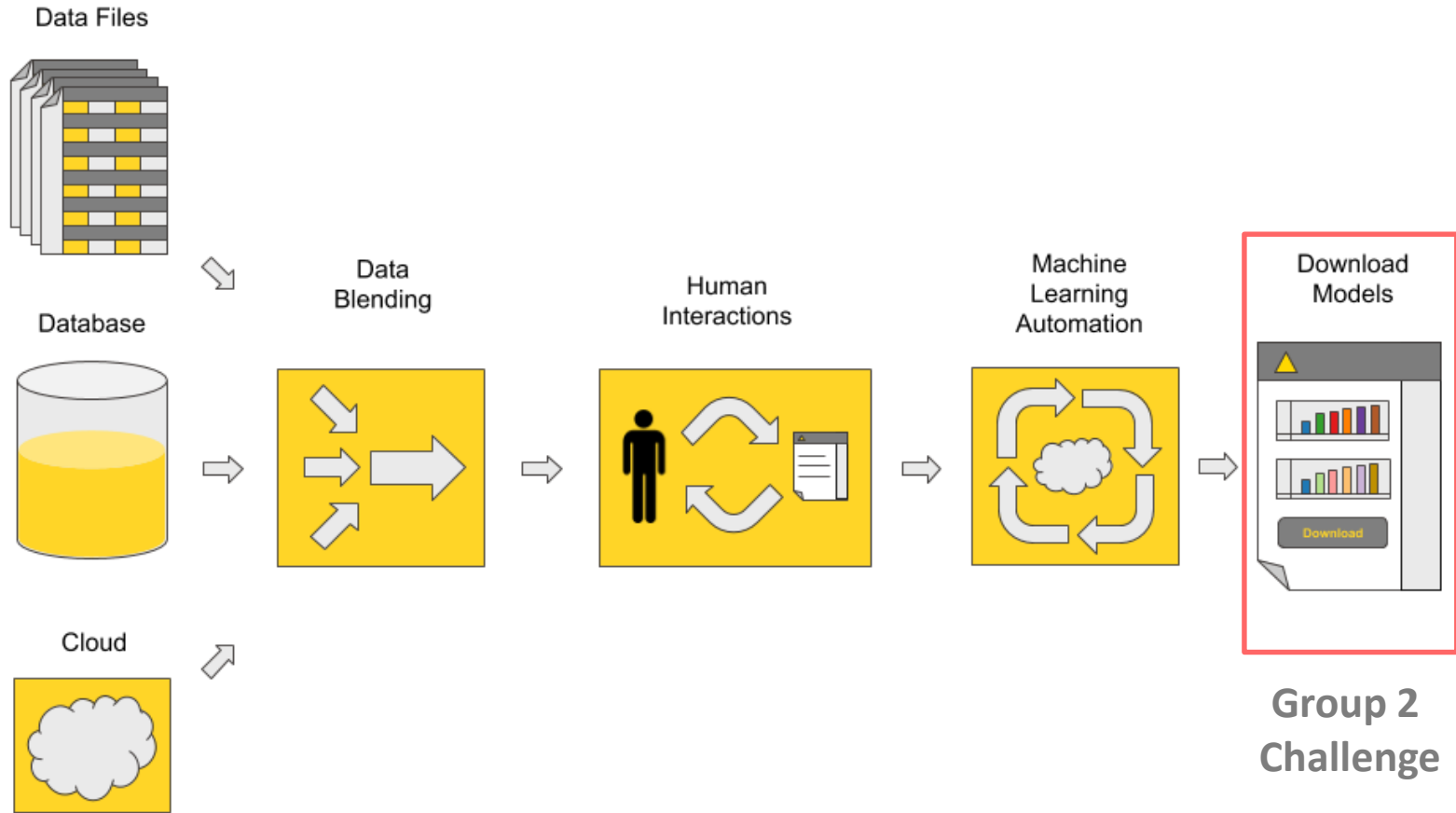


Group 1: User Interface for Automated ML Settings

- Read and inspect the data
- Define target variable
- Select which ML models to train
- Exclude certain columns from model training
- Create an interactive composite view **[OPTIONAL]**

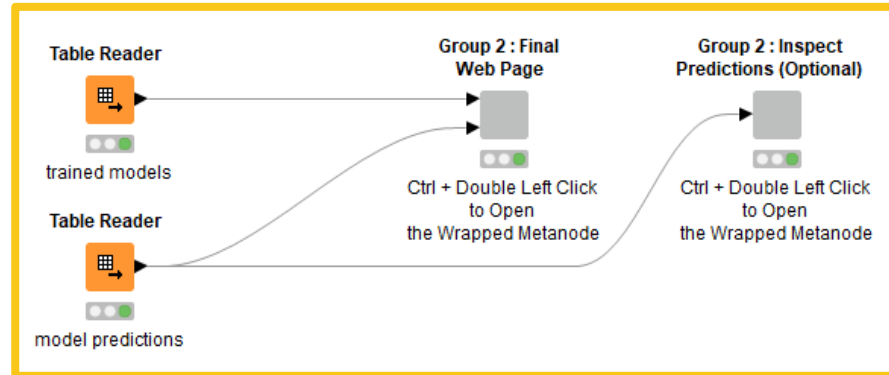
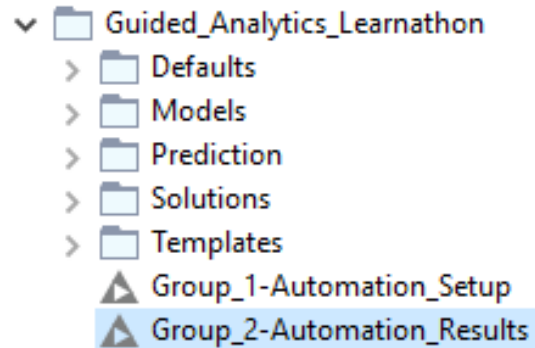


Group 2: Dashboard with Models Performance



Group 2: Dashboard with Models Performance

- Visualize and compare models by:
 - accuracies of models
 - training and scoring time
- Create buttons for model and prediction download
- Create an dashboard to inspect predictions **[OPTIONAL]**







Time to import the workflows from the USB sticks!

From the USB stick or from tinyurl.com/ODSC-Learnathon import file ***Guided_Analytics_Learnathon.knar*** into your workspace:

1. File > Import KNIME Workflows..
2. “Browse” in the *Select File* field
3. Select *Guided_Analytics_Learnathon.knar* from the USB stick or where ever else you saved it

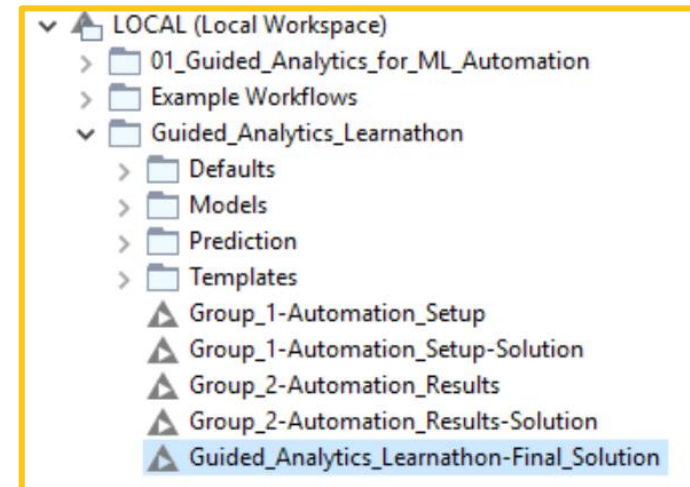
Installing Required Extensions..

- **OPTION 1:** File > Install KNIME Extensions..

<input checked="" type="checkbox"/>	 KNIME Data Generation	3.7.0.v201808081048
<input checked="" type="checkbox"/>	 KNIME H2O Machine Learning Integration	3.7.2.v201904170930
<input checked="" type="checkbox"/>	 KNIME JavaScript Views (Labs)	3.7.2.v201904170930
<input checked="" type="checkbox"/>	 KNIME PMML Preprocessing Applier Nodes	3.7.0.v201808081048

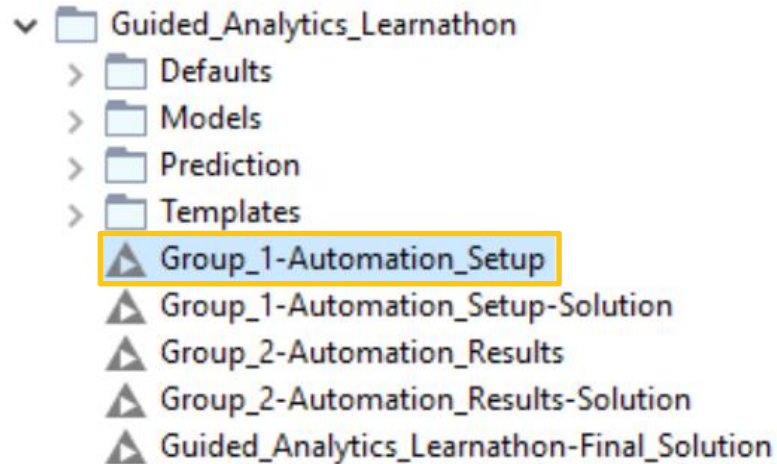
OR..

- **OPTION 2:** Just open the solution workflow and follow the instructions..

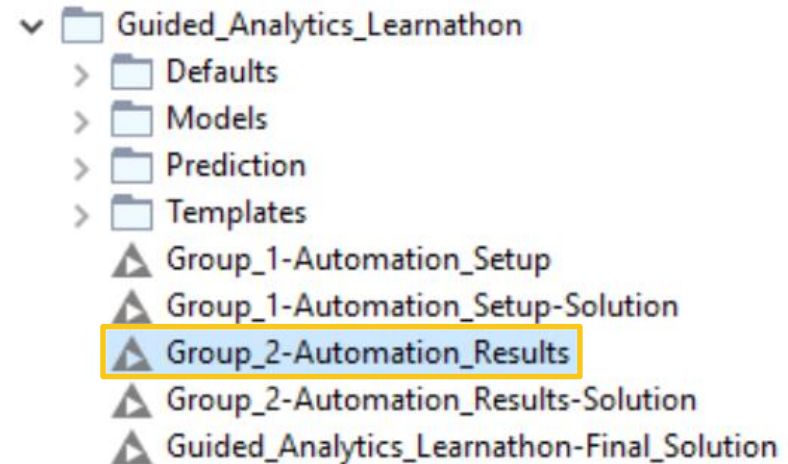


Pick your group!

- **Group 1.**
*User Interface for
Automated ML Settings*



- **Group 2.**
*Dashboard with Models
Performance*

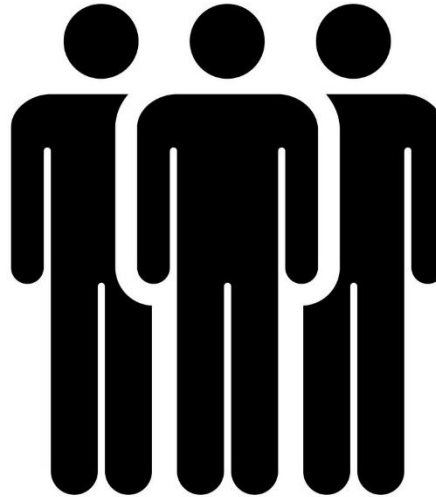


Ideally...



Data Scientist

That's it! Perfect!
Here we go!



Business Analysts



User Interactions

Business analysts will simply access the *KNIME WebPortal* from any web browser..

1 Upload Dataset
Upload the dataset to be used.
Uploading file "adult.csv"

2 Select Target
Select the target column whose values should be predicted.
Select: Label

Row ID	Workclass	Education	Education-Num	Marital St
Row0	State-gov	Bachelors	13	Never-mar

3 Filter Columns
Set Column Relevance Filter
Use the slider to select a subset of columns based on their relevance. If in doubt, do not change.
4.55
0.00 100.00
Overall Column Relevance

Feature Name	Overall Column Relevance	Correlation with Target (%)	ID/Noise Test (%)	Constant Value Test (%)	Missing Value Test (%)
Age	97.13	33.707	0.7	2.87	0
Occupation	87.12	35.595	0.14	12.88	0

4 Select Models
Choose one or more machine learning models to train for your prediction task.
Simple models
☒ Naive Bayes
☒ Decision Tree
☒ Logistic Regression
Complex models
☐ Support Vector Machine
☒ Random Forest
☒ Generalized Linear Models
☒ Gradient Boosted Trees
☒ Deep Learning

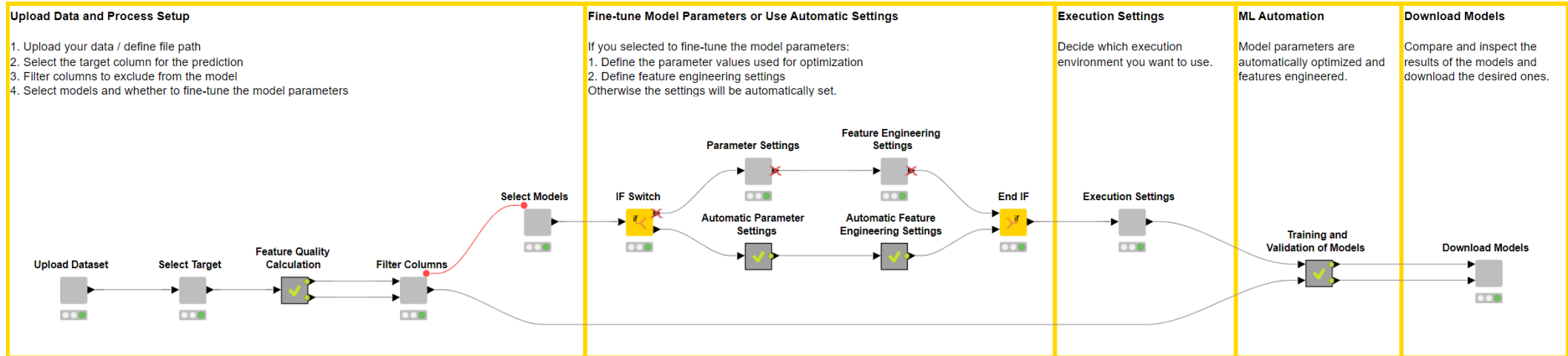
5 Execution Settings
Please select the desired distributed environments for the execution of the workflow.
Available options:
☒ Local execution
☐ Use Spark cluster if possible
☐ Use Apache Spark MLlib
☐ Use other cluster environment

Guide
Set Column Relevance Filter
By default, all columns will be used to train the model that creates the prediction. However, not all columns contribute with the same importance or relevance to the final prediction. In some cases, columns are not informative or contain spurious information. To help you decide, the overall column relevance towards the final prediction is measured.
• **Column Relevance** is an overall metric summarizing the metrics below. Use the slider to select the input features based on their Overall Column Relevance.
The additional metrics calculated automatically and used to determine Overall Column Relevance include:
• **ID/Noise Test** measures how likely the column is a representation used to identify each row in your table. Row identifiers are uninformative for your model and should be removed.
• **Constant Value Test** measures how often the column contains the exact same value. Columns with such a constant value often

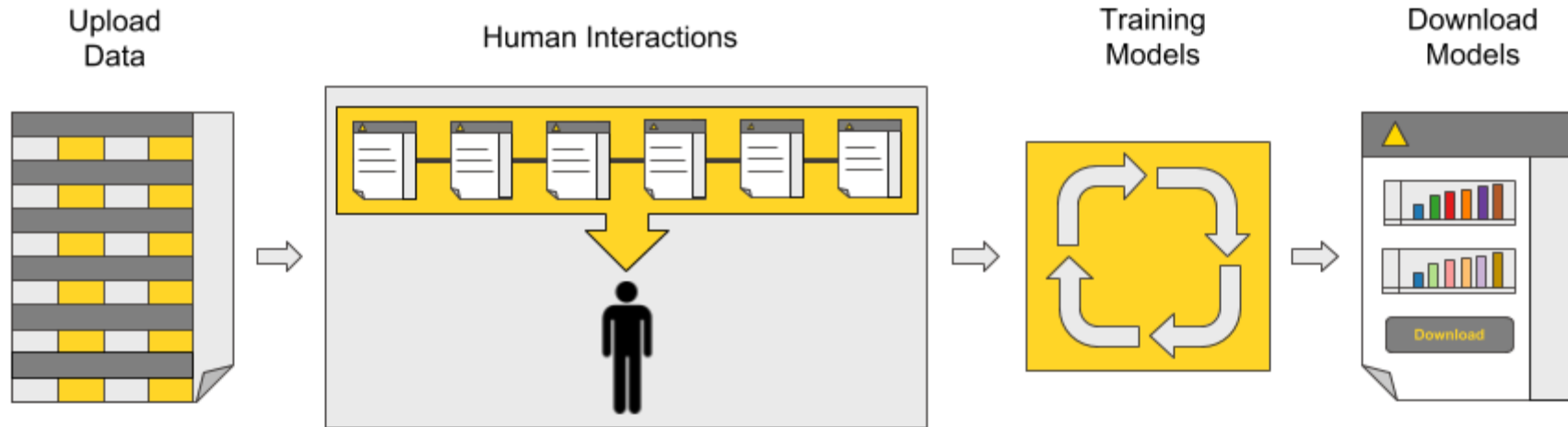
The Workflow

Guided Automation

This workflow defines a fully automated web based application to select, train, test, and optimize a number of machine learning models. The workflow was designed for business analysts to easily create predictive analytics solutions by applying their domain knowledge. Each of the wrapped metanodes outputs a web page with which the business analyst can interact.



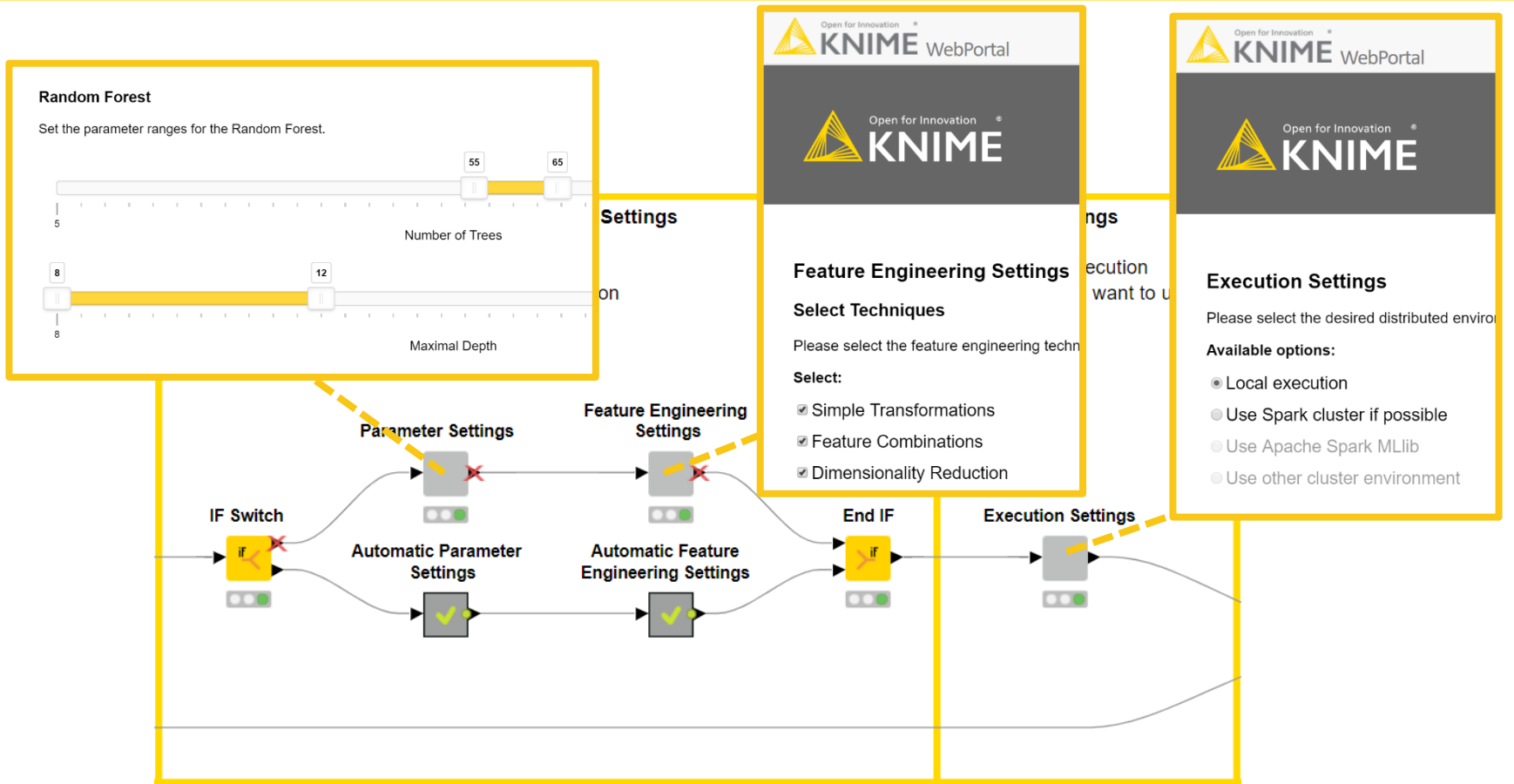
A Blueprint for Guided Automation



The Workflow



The Workflow



The Workflow

Open for Innovation
KNIME WebPortal

Open for Innovation
KNIME

Upload Dataset

Select Target

Filter Columns

Select Models

Parameter Settings

Feature Eng. Settings

Execution Settings

Download Models

Download Models

Here is a summary of information (performances) about the models trained based on your specifications. The first chart compares the accuracy and Area under the Curve of each model. The second chart compares the training times. The third chart compares the prediction time on a new record. The fourth chart shows the ROC (AUC). After the table to download the model parameters, a performance summary for each model is shown.

Compare Model Metric Performance

This bar chart visualize different performance metrics to assess the quality of each model.

Main Performance Metrics

Area Under Curve (%) Accuracy (%) zoom

Model	Area Under Curve (%)	Accuracy (%)
Naive Bayes	90	80
Generalized Linear Model	90	80
Decision Tree	85	80
Deep Learning	90	80
Logistic Regression	50	80
Gradient Boosted Trees	90	80
Random Forest	90	80

REST API

Guide

Download Models

The models shown specifications.

Each model has its feature engineering either the automatic. By means of the vis selected models to

The first chart show higher accuracy is the amount of time need chart. If a model is a amount of time may needs to be re-train then the training time chart shows the rela the model to (or sco have many records amount of time and this time is importan

The fourth chart sho additional way of lo

The information pro help you decide whi model with the high longer to be applied less accurate mode execute faster. The depend on a combin

Details for Expe

You can compare th metrics. Above you measures of perform percentage of corre predictions. **Area U** area under the **Rec** (ROC) curve.

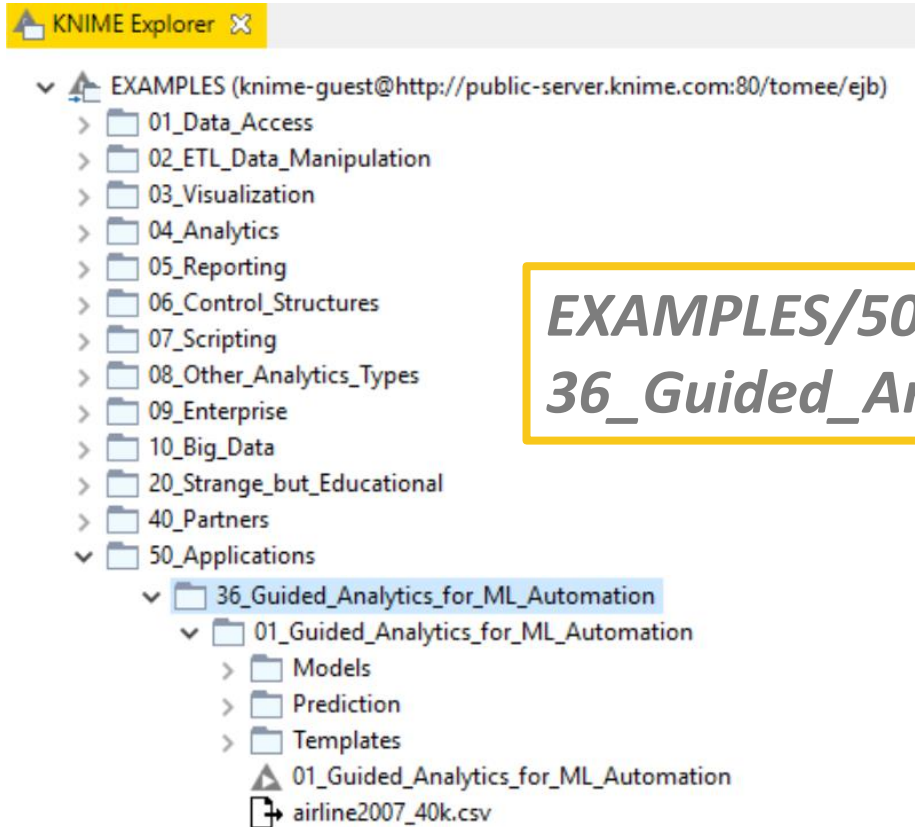
After that, if enabled, parameters for the models

Retrain the optimized models and collect results.

Output:
1. Models
Training data
(2) data (with features)


Output:
1. Models
2. Predictions - testing data

The Workflow



*EXAMPLES/50_Applications/
36_Guided_Analytics_for_ML_Automation*

On the Workflow Hub..



Search workflows, nodes and more...

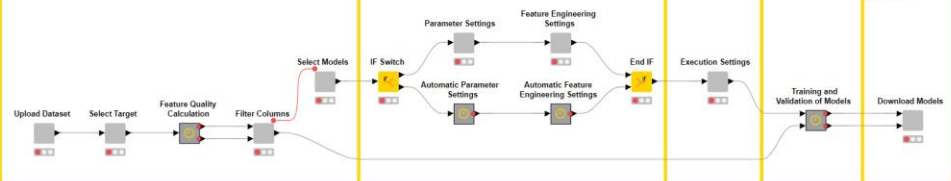
New to KNIME? More KNIME

Guided Analytics for Machine Learning Automation


Guided Automation

This workflow defines a fully automated web based application to select, train, test, and optimize a number of machine learning models. The workflow was designed for business analysts to easily create predictive analytics solutions by applying their domain knowledge. Each of the wrapped metanodes outputs a web page with which the business analyst can interact.

Upload Data and Process Setup	Fine-tune Model Parameters or Use Automatic Settings	Execution Settings	ML Automation	Download Models
<ol style="list-style-type: none">1. Upload your data / define file path.2. Select the target column for the prediction.3. Filter columns to exclude from the model.4. Select models and whether to fine-tune the model parameters.	<p>If you selected to fine-tune the model parameters:</p> <ol style="list-style-type: none">1. Define the parameter values used for optimization.2. Define feature engineering settings. <p>Otherwise the settings will be automatically set.</p>	<p>Decide which execution environment you want to use.</p>	<p>Model parameters are automatically optimized and features engineered.</p>	<p>Compare and inspect the results of the models and download the desired ones.</p>




Upload Dataset → Select Target → Feature Quality Calculation → Filter Columns → Select Models → IF Switch → Parameter Settings / Automatic Parameter Settings → Feature Engineering Settings / Automatic Feature Engineering Settings → End IF → Execution Settings → Training and Validation of Models → Download Models



Simon Schmid
KNIME Team Member
Software developer and data scientist at KNIME.

Leader Silver Anniversary
Thoughtful 12 more

hosted by
 **KNIME**

[Open workflow](#)
or [download workflow](#)

By downloading the workflow, you agree to our [terms and conditions](#).

License [CC-BY-4.0](#)

tinyurl.com/GuidedAutomation

KNIME Fall Summit 2019

November 5 – 8 at AT&T Executive Education and Conference Center, Austin, Texas

- **Tuesday & Wednesday:** One-day courses
- **Thursday & Friday:** Summit sessions

10% off Promo Code:

ODSC-EAST

Register at
knime.com/fall-summit2019



KNIME Beginner's Luck

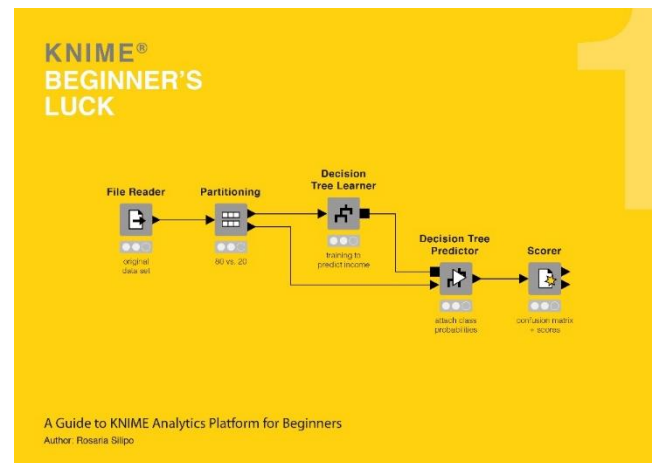
Free Copy of KNIME Beginner's Luck Book from **KNIME Press**

knime.com/knimepress





with this code: **ODSC-BOSTON-2019**



#KNIME
#ODSCEast



Keep in Touch with KNIME!

- **KNIME** (www.knime.com)
 - **BLOG** for news, tips and tricks(www.knime.com/blog)
 - **FORUM** for questions and answers (tech.knime.com/forum)
 - **EXAMPLE SERVER** for example workflows
 - **LEARNING HUB** (www.knime.com/learning-hub)
- **KNIME TV** channel on 
- **KNIME** on  **@KNIME**
- **KNIME** on  <https://www.facebook.com/KNIMEanalytics>
- On 

Thank You!



#KNIME
#ODSCEast

The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.