



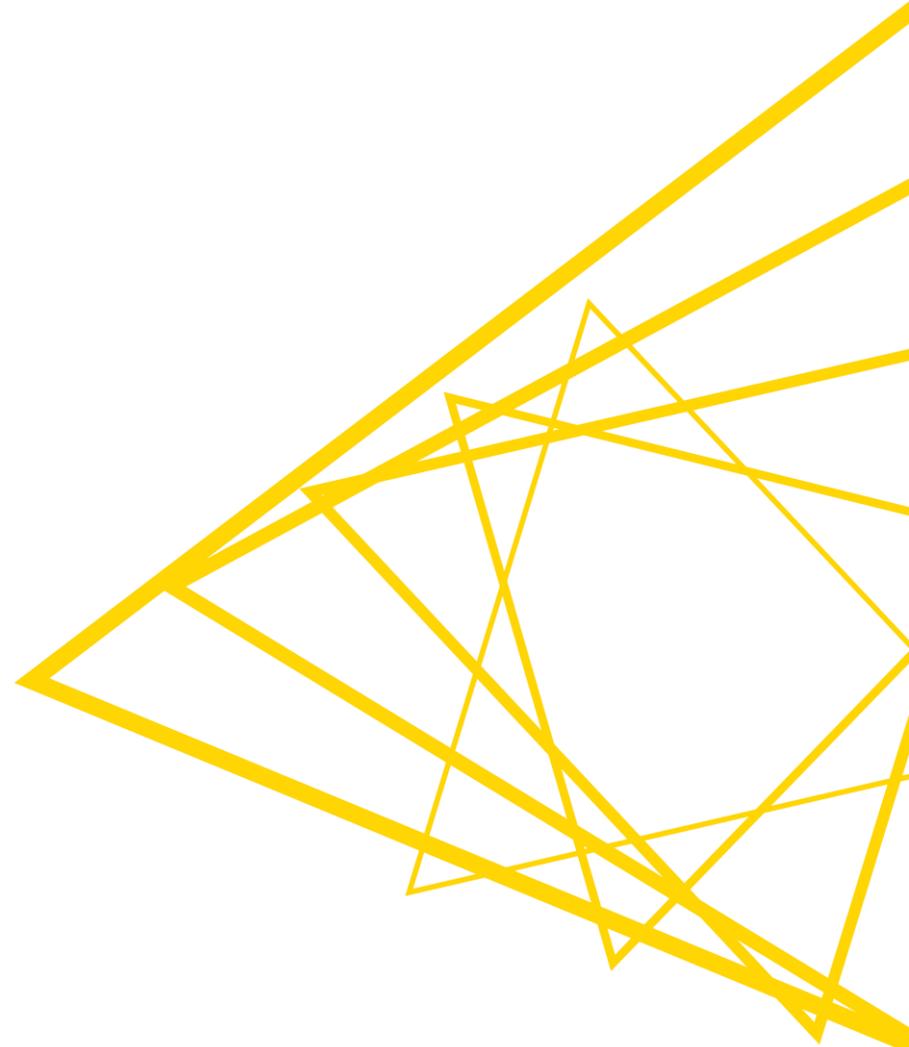
Open for Innovation

KNIME

Data Science in Action: Stakeholder Collaboration

Paolo Tamagnini
Phil Winters

& the KNIME Team



Collaboration

Two or more people working together to complete a task or achieve a goal

Structured



Unstructured



Stakeholders involved in Data Science.....



Team Leader
Data Science Practice



Data Engineers
Data Scientists
“Code First” / Package First



Business Analyst
Interactive User
Information Consumer



Model / ML Operations



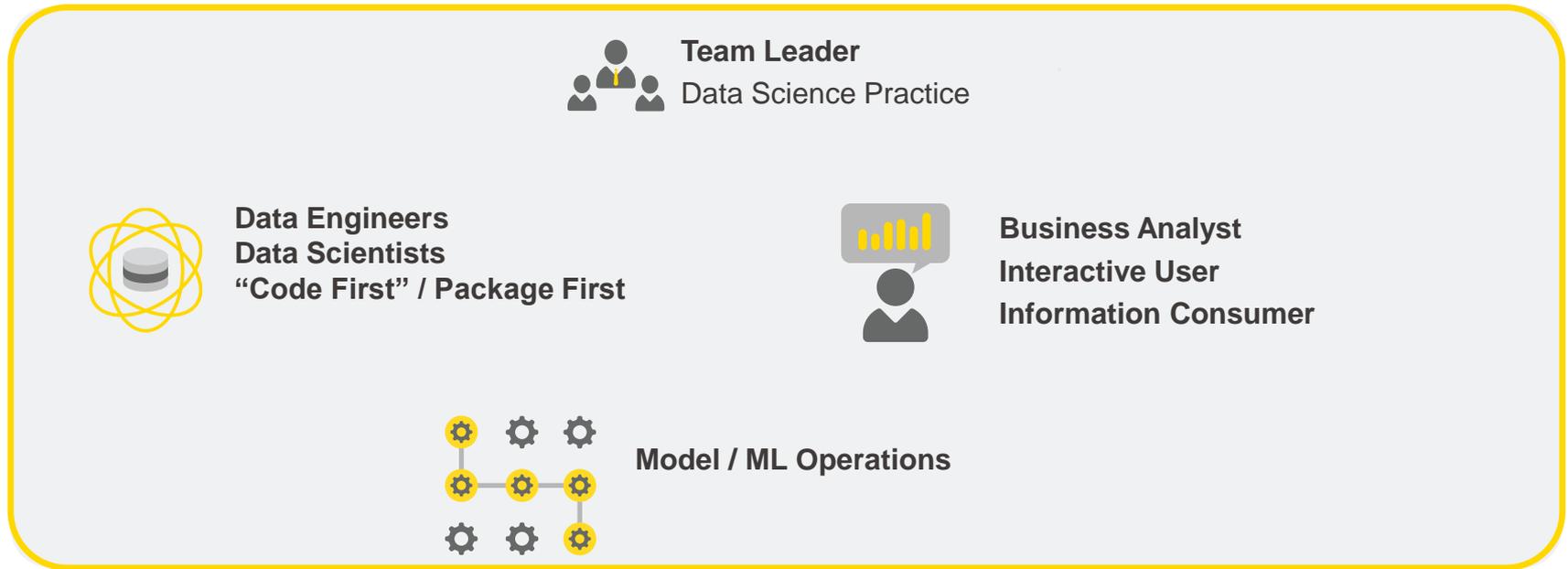
IT Operations
Scalability and control
Centralized resources / strategies
Security



Compliance
Financial / risk oversight
Costs allocation
Governance, traceability, GDPR

Successful Team Collaboration.....

Users



IT Operations
Scalability and control
Centralized resources / strategies
Security

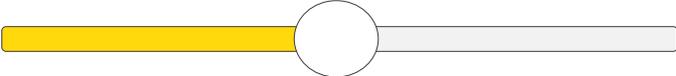
Compliance
Financial / risk oversight
Costs allocation
Governance, traceability, GDPR

Common across ALL Users, just different emphasis

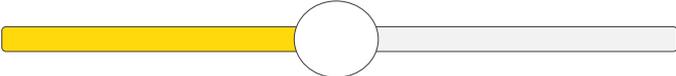
Low Focus

High Focus

Create Great Results



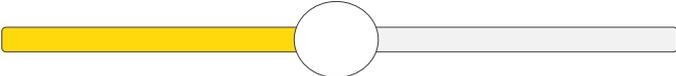
Use Machine Learning



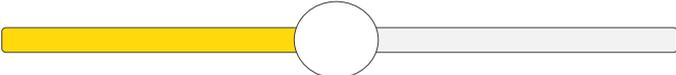
Deploy & Automate



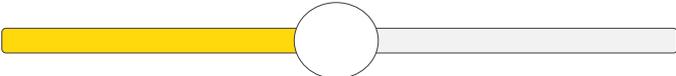
Understand the Data



Understand the Business



Responsible for the Business



Data Engineers / Data Scientists



Phil



Kathrin

SAME, only a difference in ML Focus!

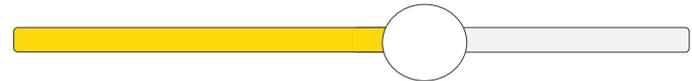
Create Great Results



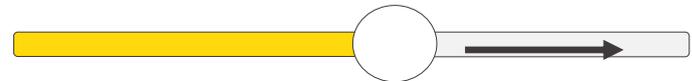
Use Machine Learning



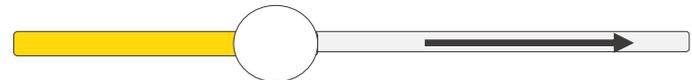
Deploy & Automate



Understand the Data



Understand the Business

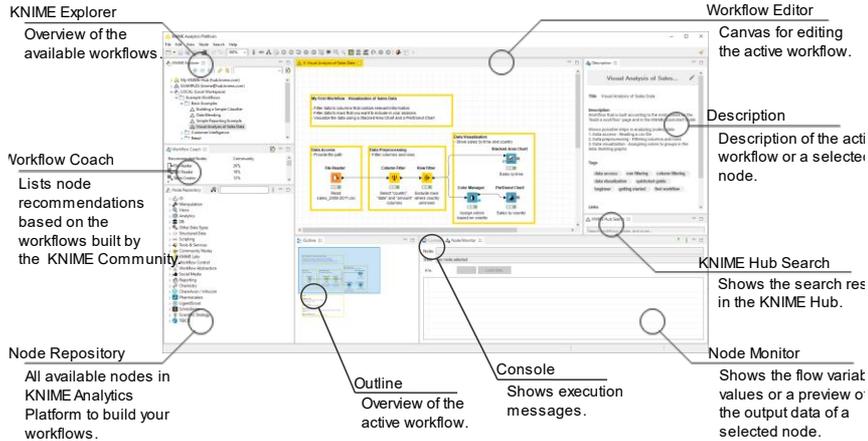


Responsible for the Business

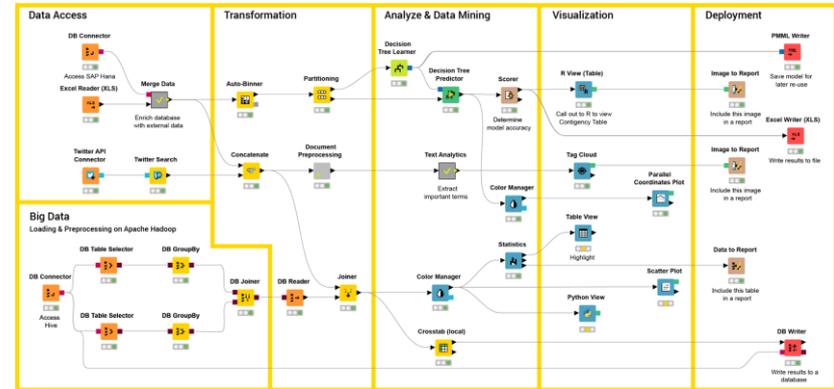


KNIME Provides

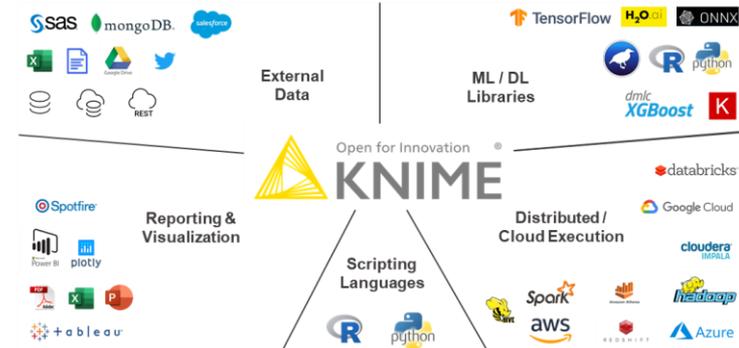
KNIME Workbench



3500+ Nodes



100+ extensions



Data Engineers / Data Scientists / “Code-First”



Phil



Kathrin



Monty

No Difference! Only preference
(Python in this case !)

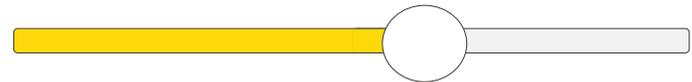
Create Great Results



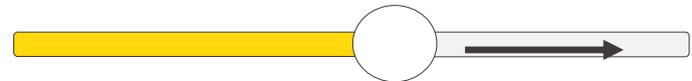
Use Machine Learning



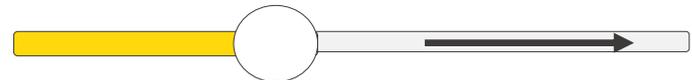
Deploy & Automate



Understand the Data



Understand the Business

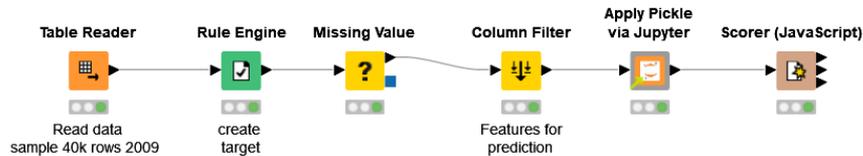


Responsible for the Business



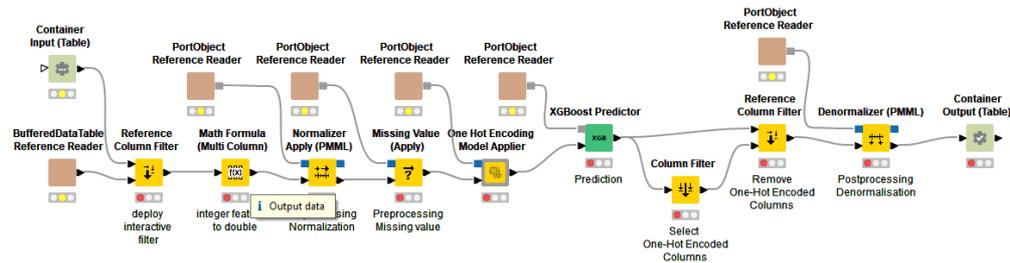
KNIME Provides: IDEs and Code-First Specifics

- Using Jupyter Notebooks in KNIME



- Using KNIME in Jupyter Notebooks

```
In [1]: import pandas as pd
import knime
knime.executable_path = "c:/Users/glandrum/Desktop/KNIME_3.7/knime.exe"
```



Model / ML Operations



Ana

Create Great Results

Use Machine Learning

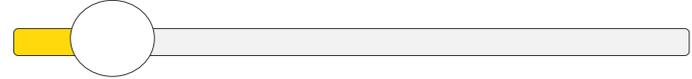
Deploy & Automate

Understand the Data

Understand the Business

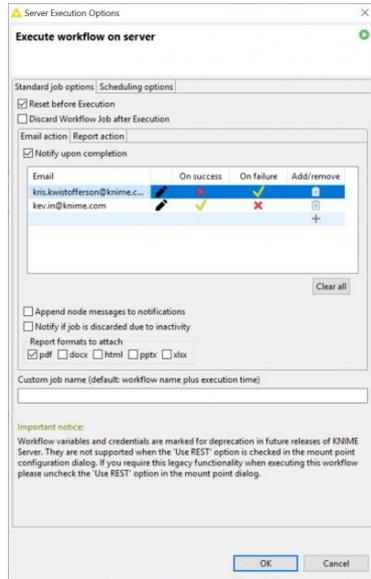
Responsible for the Business

Deploy, otherwise it is useless...



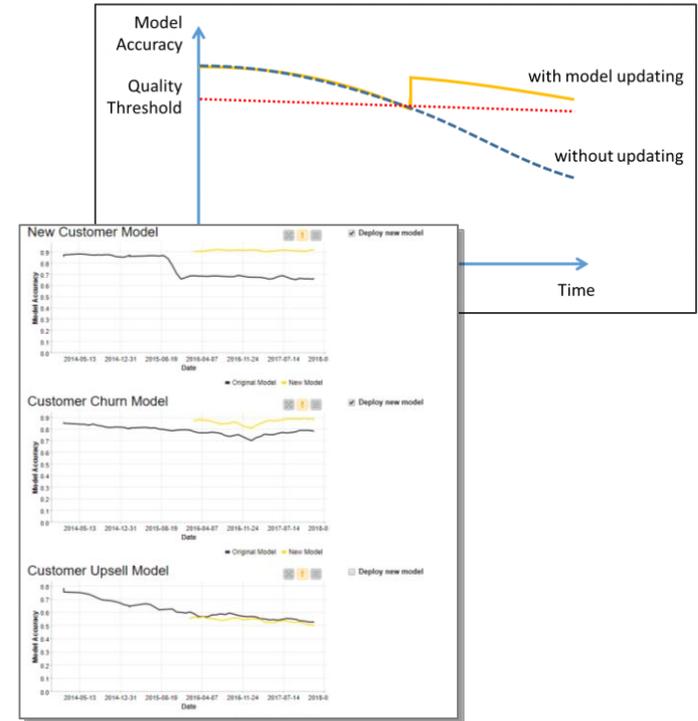
KNIME Provides: Automation Options

Automated workflow execution



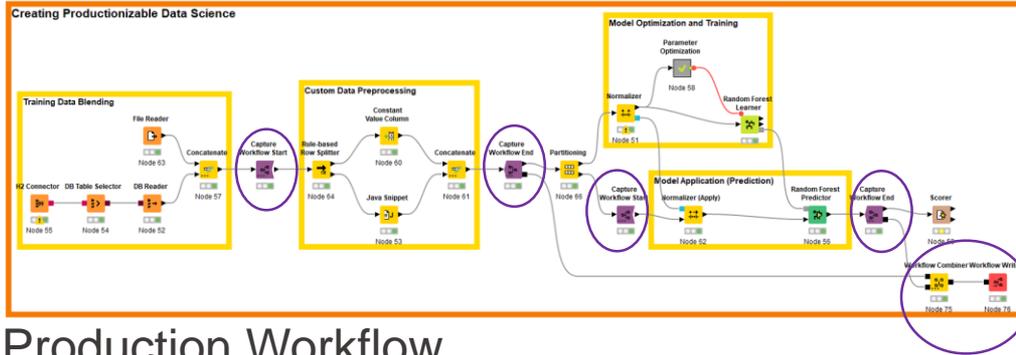
- Scheduled
- Triggered/Manual Model Called
- (Rest / SAAS)
- Champion / Challenger
- Model Monitoring
- Retraining
- View, edit, execute workflows remotely

KNIME Model Process Factory

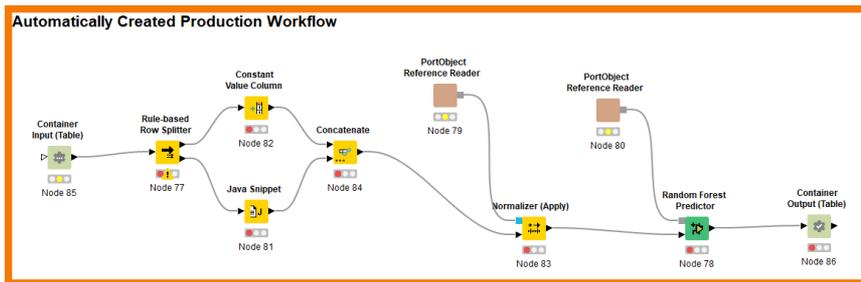


KNIME Provides: Integrated Deployment

Creation Workflow



Production Workflow



- Eliminates the gap between Creation and Production
 - mark what's necessary for production
 - not just the model
- no manual intervention
 - No rewrite, copy paste
- Always in sync

Successful Team Collaboration.....



Team Leader
Data Science Practice



Data Engineers
Data Scientists
“Code First” / Package First



Business Analyst
Interactive User
Information Consumer



Model / ML Operations



IT Operations
Scalability and control
Centralized resources / strategies
Security



Compliance
Financial / risk oversight
Costs allocation
Governance, traceability, GDPR

Information Consumer / Interaction User



Heather



Christiane

Create Great Results

Use Machine Learning

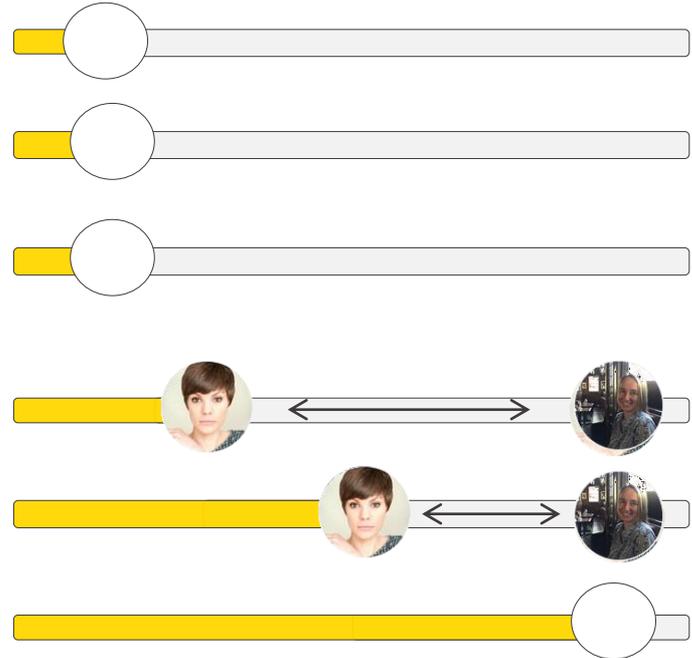
Deploy & Automate

Understand the Data

Understand the Business

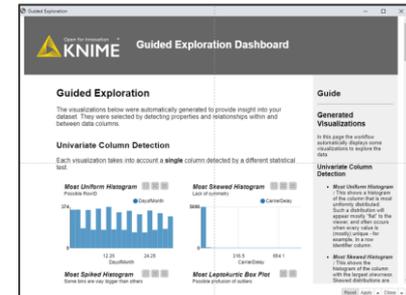
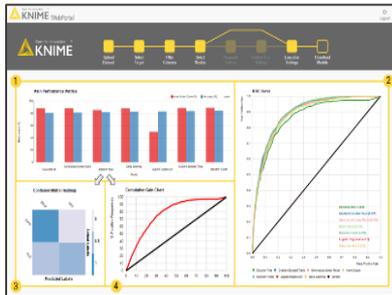
Responsible for the Business

Use workflows.....but never see them.

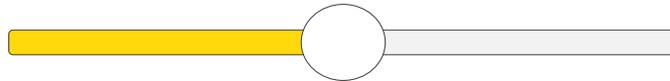


KNIME Provides: Guided Analytics

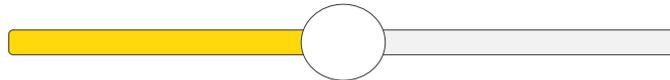
- Appropriate levels of Automation & Human Interaction depending on **task** and **audience**



Automation



Interaction

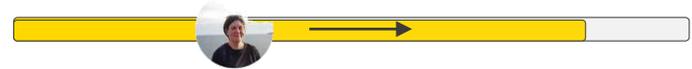


Business Analyst



Daria

Create Great Results



Use Machine Learning



Deploy & Automate



Understand the Data



Understand the Business



Responsible for the Business



KNIME Provides: Tailored KNIME Analytics Platform

Surface capabilities so users can appropriately create workflows themselves

The screenshot displays the KNIME Analytics Platform interface. On the left, a node catalog lists various nodes such as File Reader, CSV Reader, Table Creator, and Database Reader, with their respective usage percentages. The central workspace shows a workflow canvas with a 'File Reader' node connected to a 'Choropleth World Map' node. The right-hand side features a help panel for the 'Choropleth World Map' node, providing detailed instructions and options for its configuration. The interface is annotated with yellow boxes highlighting key areas: the node catalog, the workflow canvas, the help panel, and the 'Choropleth World Map' node itself.

Tailor / Focus:

- Welcome
- Help /guidance
- Available Workflows
- Node availability
- Components & Descriptions
- Workflow coach for recommendations

DS Practice Leader



Rosaria

Leverage, Exploit, Grow, Limit Risk, Satisfy

Create Great Results



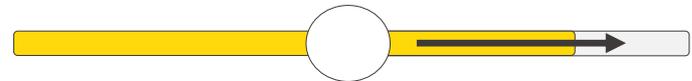
Use Machine Learning



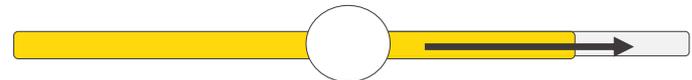
Deploy & Automate



Understand the Data



Understand the Business



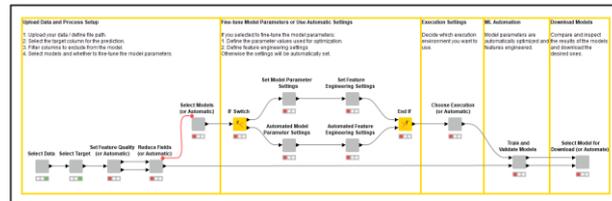
Responsible for the Business



Captured, Instantiated, Secure, Versioned, Reproducible

Workflows and Components

- knime-server-webinar (phil@http://3.80.12.216:8080/knime/rest)
 - Admin
 - Components
 - Automated Feature Engineering Settings
 - Automated Model Parameter Settings
 - Choose Execution (or Automatic)
 - Reduce Fields (or Automatic)
 - Select Data
 - Select Model for Download (or Automate)
 - Select Models (or Automatic)
 - Select Target
 - Set Feature Engineering Settings
 - Set Feature Quality (or Automatic)
 - Set Model Parameter Settings
 - Train and Validate Models
 - Data
 - 2007.csv
 - 2008.csv
 - adults.csv
 - Examples
 - Users
 - Workflows
 - 01_Guided_Analytics_for_ML_Automation.knwf 1



KNIME Provides:

- Instantiation/ Updating
- Security (create, use, etc.)
- Versioning and snapshots
 - Version comparison
- Locking / Encryption
- Reproducibility
- Guaranteed backward compatible

Snapshots for /Workflows/01_Guided_Analytics_for_ML_Automation.knwf 1

Element/Creation date	Creator	Comment
Mon 2020-06-08, 09:36:42h	phil	Reworded label text of a few of the user interaction nodes
Mon 2020-06-08, 09:34:40h	phil	Tweaked parameter optimization settings.
Mon 2020-06-08, 09:33:37h	phil	Updated documentation

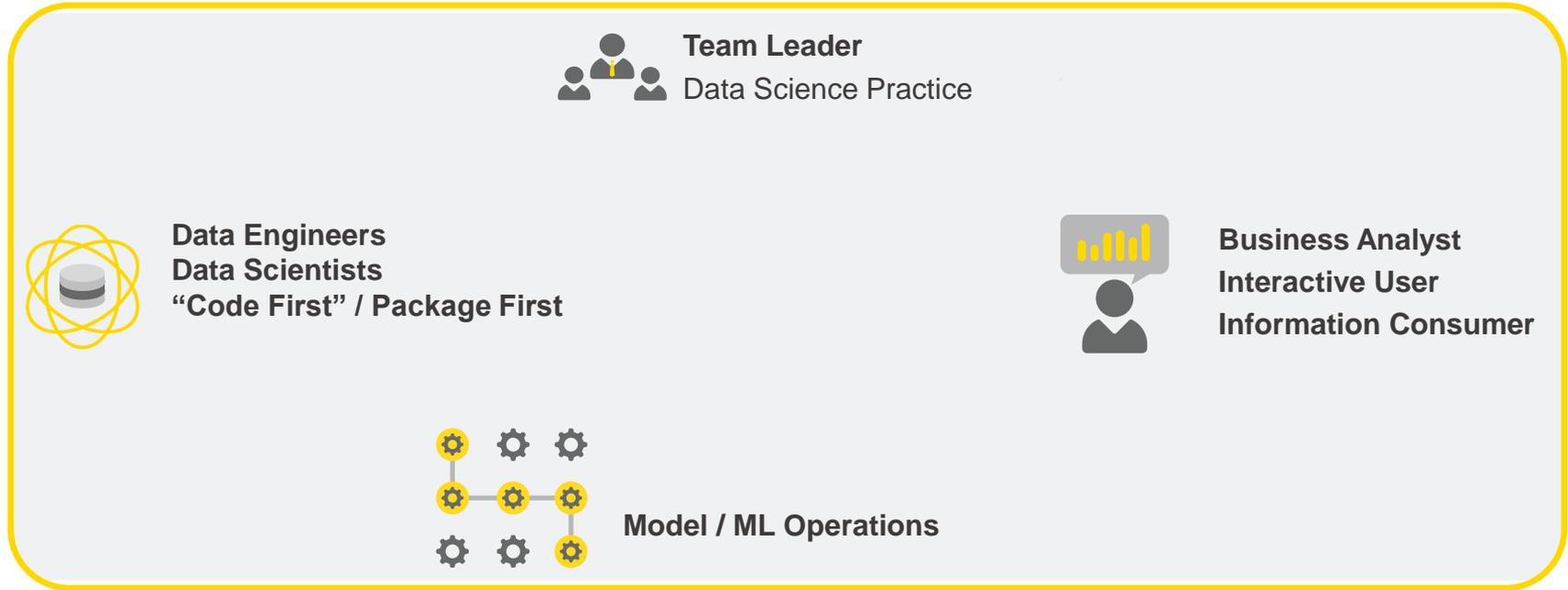
Explainability / Interpretability

KNIME Provides:

- Methods
- Examples
- Interactive Components
 - XAI



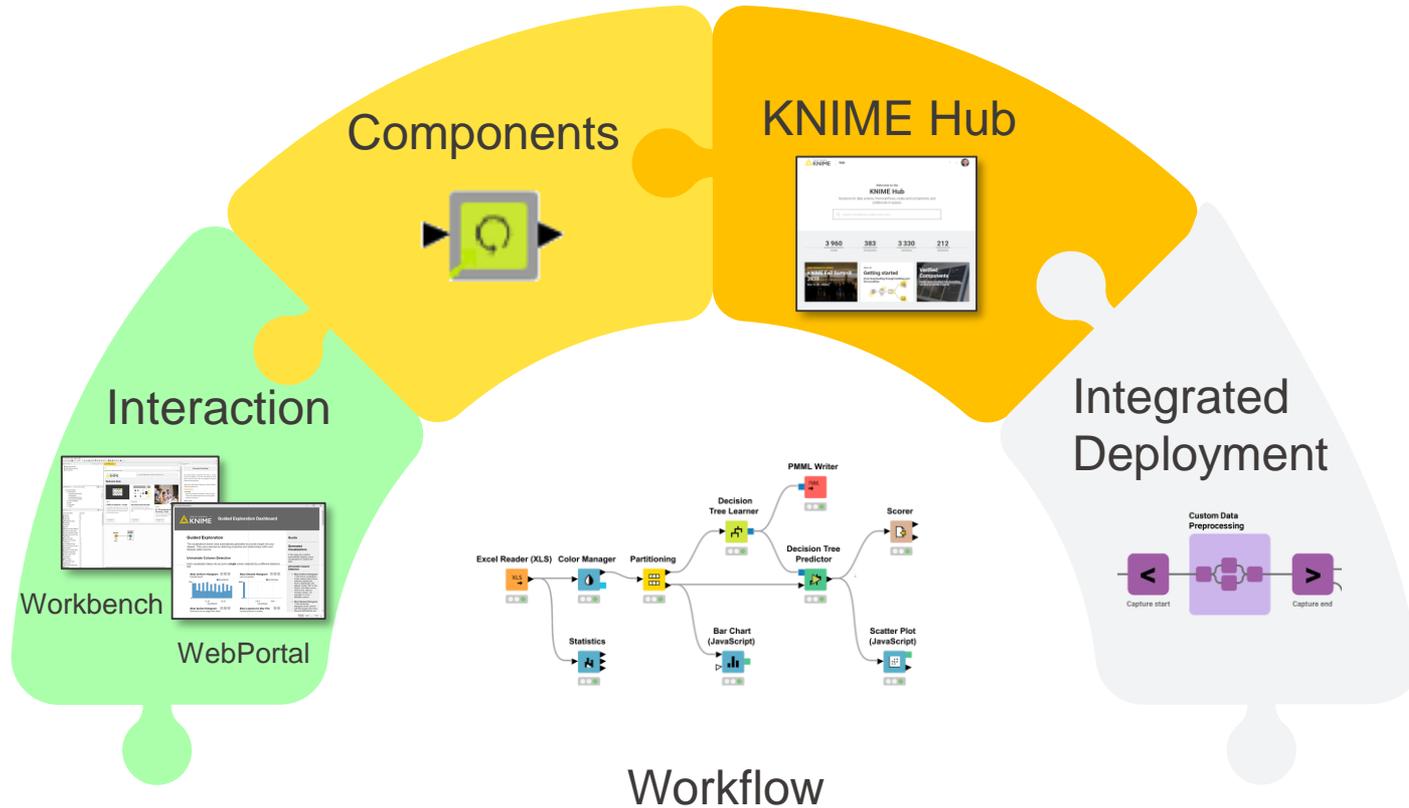
Successful Team Collaboration Uses The KNIME Platform



IT Operations
Scalability and control
Centralized resources / strategies
Security

Compliance
Financial / risk oversight
Costs allocation
Governance, traceability, GDPR

Highlights for Team Collaboration



Live Demo - Workflows

KNIME Hub

HUB.KNIME.COM

 **Public & Private Spaces**

 **Search**

 **Drag & Drop**

 **4**  **599**

 **Discussion**

 **KNIME Hub Search Bar in
KNIME Analytics Platform**

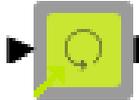
 **Workflow Metadata Editor**

.... a peak into what's coming:

**Using KNIME Hub
capabilities in the Enterprise**

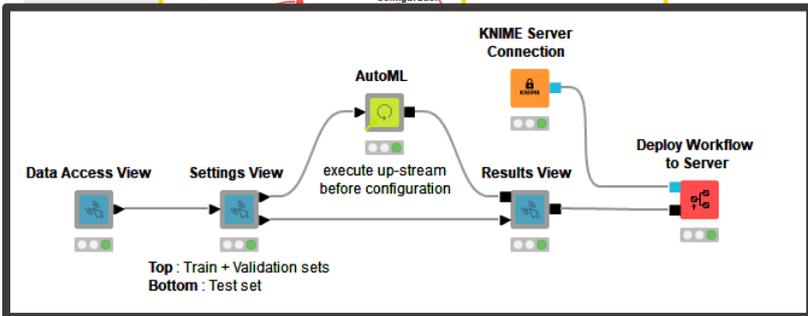
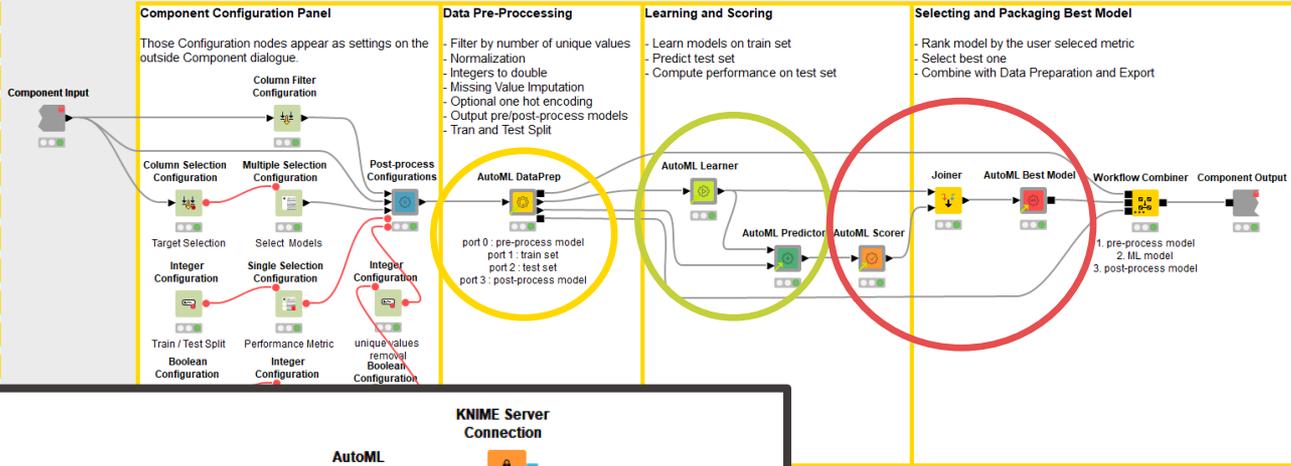
Collaboration on a complex Component

AutoML



Automated Machine Learning (AutoML)

This component can automatically train supervised machine learning models by performing some data preparation, parameter optimization, scoring, evaluation and selection.



Data Access and Preparation

Before training the models the data is cleaned by replacing the missing values with the categorical column most frequent value ...

Last update: 13 Nov 2020

0 1 0

Machine Learning

- Naive Bayes: trained with optimized parameter "Default probability".
- Logistic Regression: trained with optimized...

Last update: 13 Nov 2020

0 2 0

Deployment

The best model is selected using the performance metric specified by the user. The predictions of all models are scored...

Last update: 13 Nov 2020

0 2 0

User Testing

This application is a simple example of AutoML with KNIME Software for binary and multiclass classification. The key of this...

Last update: 13 Nov 2020

1 1 0

Collaboration on our Component



paolotamag 
KNIME Team Member
Berlin, Germany

Overview 

Extensions 

Spaces 

Your spaces



Machine Learning

- Naive Bayes: trained with optimized parameter "Default probability". - Logistic Regression: trained with optimized...

Last update: 13 Nov 2020

 0  2  0





User Testing

This application is a simple example of AutoML with KNIME Software for binary and multiclass classification. The key of this...

Last update: 13 Nov 2020

 1  1  0





Data Access and Preparation

Before training the models the data is cleaned by replacing the missing values with the categorical column most frequent value ...

Last update: 13 Nov 2020

 0  1  0





Deployment

The best model is selected using the performance metric specified by the user. The predictions of all models are scored...

Last update: 13 Nov 2020

 0  2  0

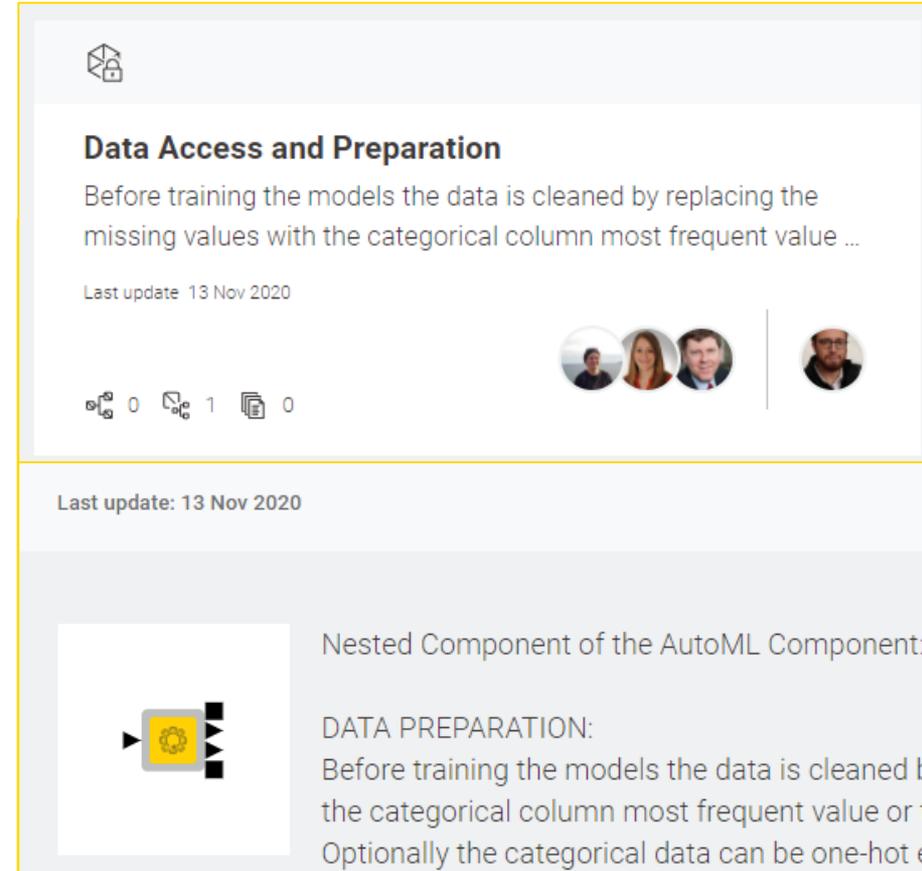
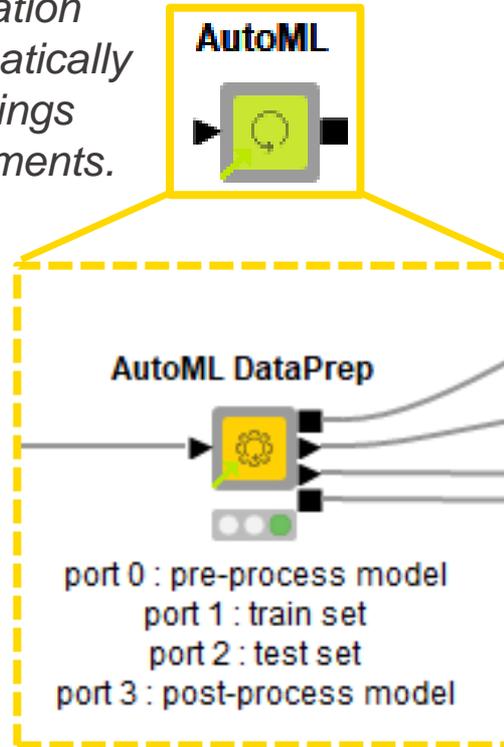


“Data Access and Preparation” Collaboration Space

A team making sure all the data preparation would place automatically based on user settings and model requirements.

Collaborators:

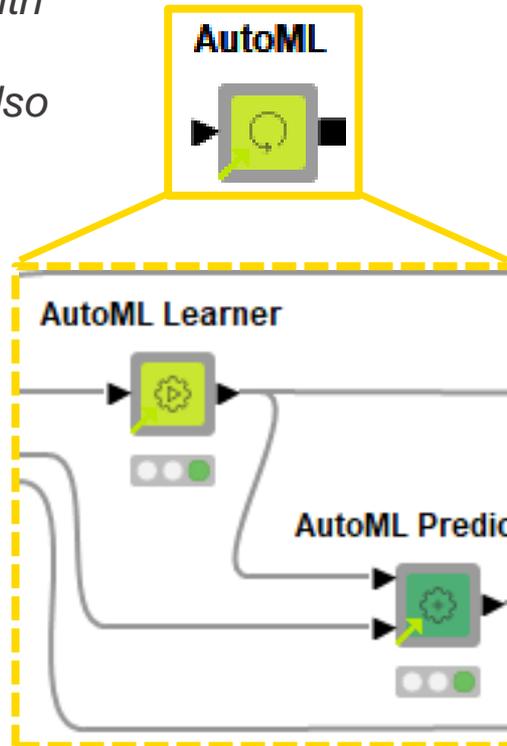
-  Project Owner
-  Data Engineer
-  Data Scientist
-  Business Analyst



The screenshot shows a collaboration space interface. At the top left is a lock icon. The main title is "Data Access and Preparation". Below the title is a description: "Before training the models the data is cleaned by replacing the missing values with the categorical column most frequent value ...". The last update is "13 Nov 2020". There are four profile icons of collaborators. Below the icons are icons for comments (0), likes (1), and shares (0). The bottom section shows a smaller version of the AutoML DataPrep icon and the text: "Nested Component of the AutoML Component: DATA PREPARATION: Before training the models the data is cleaned by the categorical column most frequent value or t. Optionally the categorical data can be one-hot e".

“Machine Learning” Collaboration Space

A team for training a variety of models with KNIME nodes and custom code, but also to be easily moved to production.

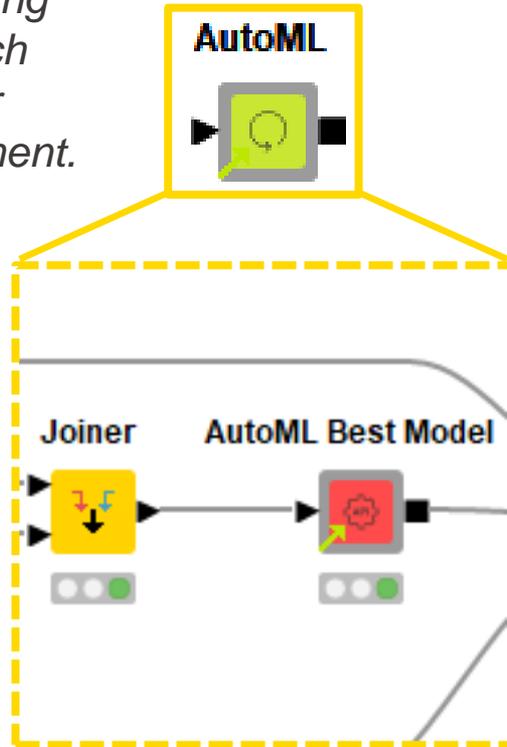


Collaborators:

-  Project Owner
-  Code First
-  Data Scientist
-  ML Ops

“Deployment” Collaboration Space

A team for selecting, building and exporting a single model which satisfies policies for the correct deployment.



Collaborators:

-  Project Owner
-  ML Ops
-  DS Practice Leader

Deployment

The best model is selected using the performance metric specified by the user. The predictions of all models are scored...

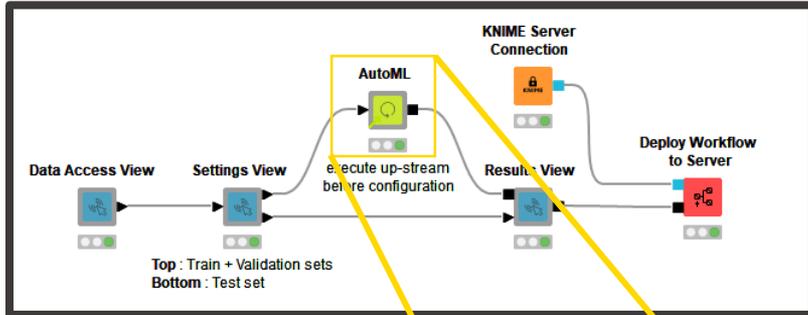
Last update: 13 Nov 2020

0 2 0

Last update: 13 Nov 2020

MODEL SELECTION:
The best model is selected using the performan

“User Testing” Collaboration Space



Collaborators:



Project Owner



Information Consumer



Business Analyst



DS Practice Leader



User Testing

This application is a simple example of AutoML with KNIME Software for binary and multiclass classification. The key of this...

Last update: 13 Nov 2020

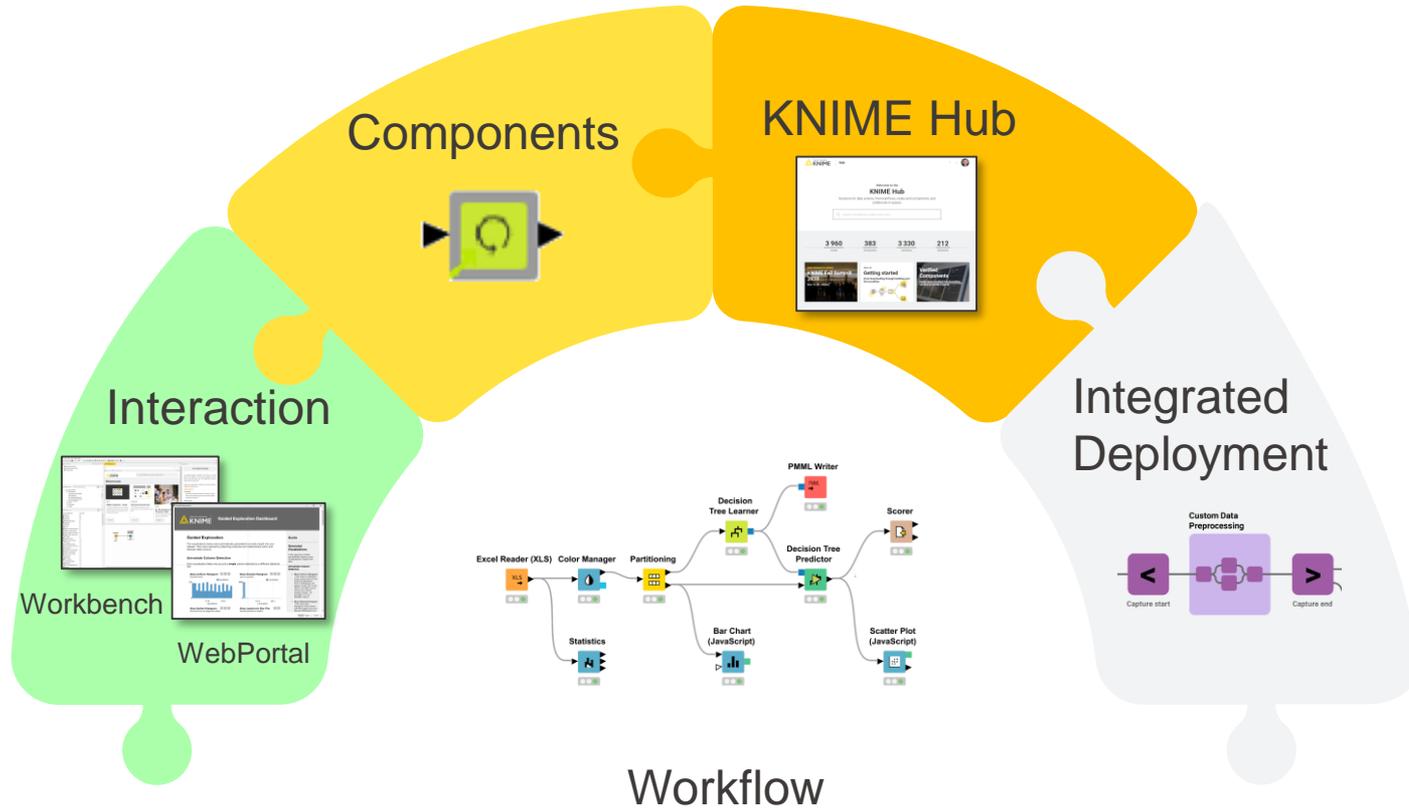
1 1 0

Last update: 13 Nov 2020

This Component automatically trains supervised binary and multiclass classification. The component completes the ML cycle by performing some data preparation, pre-validation, scoring, evaluation and selection. The component runs the end to end process and outputs the deployment configuration. Integrated Deployment Extension.

Live Demo – Collaboration

Highlights for Team Collaboration



New KNIME Pros Learnathon Tomorrow!



Group 1
Financial Analysis





Maarit **Lada**
KNIME Team Member KNIME Team Member

Group 2
Life Sciences





temesgen-dadi **Francosinus**
KNIME Team Member KNIME Team Member

Group 3
Automation





paolotamag **Mpattadkal**
KNIME Team Member KNIME Team Member

Successful Team Collaboration



Team Leader
Data Science Practice



Data Engineers
Data Scientists
"Code First" / Package First



Business Analyst
Interactive User
Information Consumer



Model / ML Operations



IT Operations
Scalability and control
Centralized resources / strategies
Security



Compliance
Financial / risk oversight
Costs allocation
Governance, traceability, GDPR

Compliance*: Governance, Auditing, GDPR (Best Practices!)

Metadata: External Documentation; Data / Model Lineage, Asset identification

Workflow Summary
Summary Created: 2020-06-25 15:38:27 +0200
Workflow Name: 02_Learning_a_Random_Forest_Phil
Description: [Redacted]
Workflow Annotations: [Redacted]

Author	Creation Date	Last Edited	Last Uploaded
[Redacted]	[Redacted]	[Redacted]	[Redacted]

Workflow Location: C:\Users\Phil\Documents\KNIME Workflows\slightly\data Lineage testing\example transformations

KNIME Version: 4.2.0.v202006250728
Operating System: Windows 10

Node, Component, Metanode Summary	NodeName	Node Type	Occurrence In Workflow
Data Explorer	Visualizer	1	
File Reader	Source	1	
Partitioning	Manipulator	1	
Random Forest Learner	Learner	1	
Random Forest Predictor	Predictor	1	
Scorer (deprecated)	Other	2	

Data Locations

NodeFullName	Node Type	Location
File Reader:1	Source	knime://knime.workflow/data/iris.data

Library, Plugins and Extension Details

plugin Name	version
org.eclipse.osgi	3.15.200.v20200214-1600
org.eclipse.osgi.compatibility.state	1.1.700.v20200207-2156
javax.transaction	1.1.1.v201105210645
org.eclipse.equinox.simpleconfigurator	1.3.500.v20200211-1505

Created with KNIME® Analytics Platform | www.knime.com

Selection

- Trace Data Source
- Trace Data Sink
- Trace Column From Source
- Identify Deprecated
- Identify Code-First
- Identify Components
- Identify Node States

Trace Data Source
C:\Users\Phil\Documents\KNIME Workflows\DMSL Metadata Exploitation\Workflow2Network\1\data\TestFlow.xml

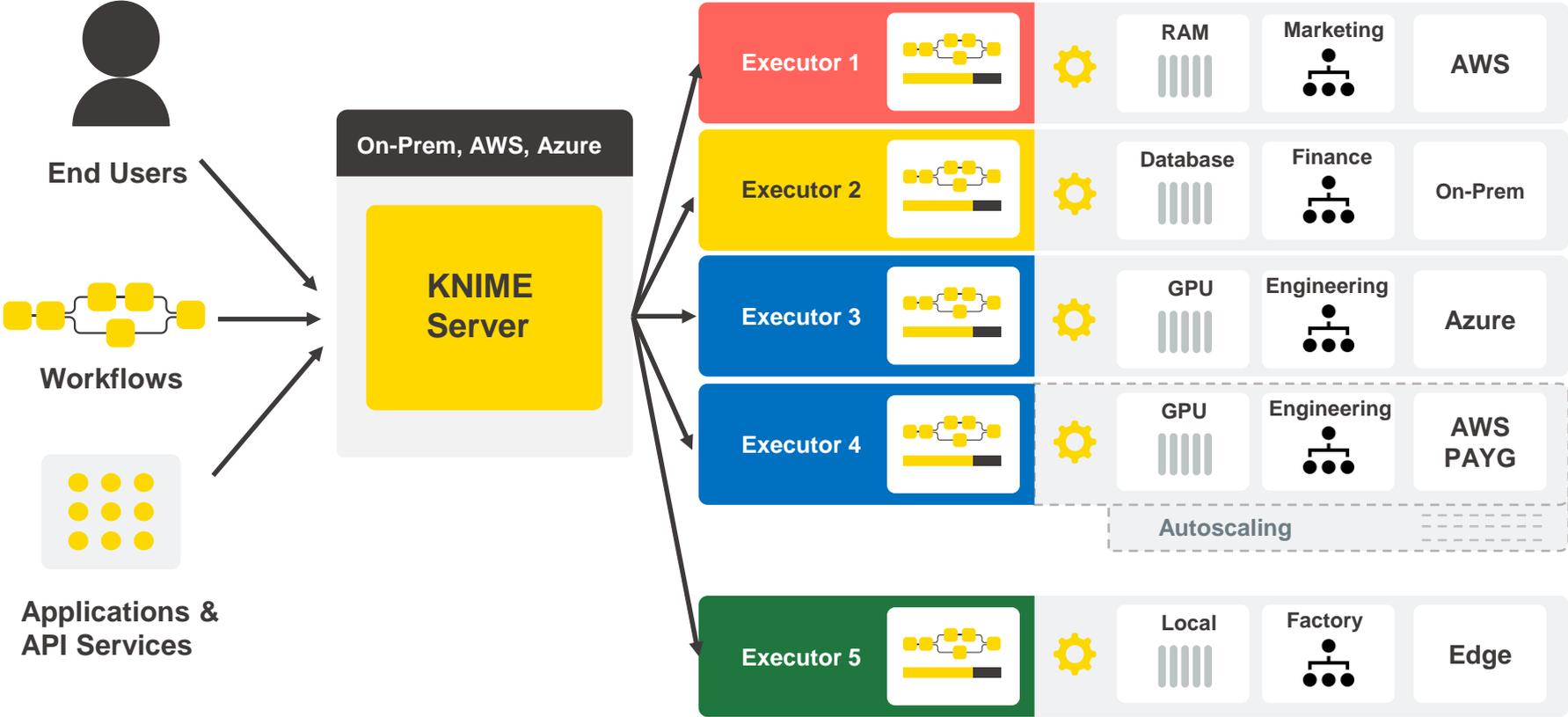
Trace Diagram

Trace Value

- Table Reader: knime://knime.workflow/data/wine.table
- Table Reader: knime://knime.workflow/data/wine2.table
- H2 Connector: temporary
- Model Reader: C:\Users\Phil\Documents\KNIME Workflows\KN

*Don't tell them, but compliance folks are just Information Consumers / Interaction Users!

IT Operations: Choice, control and definable costs



Successful Team Collaboration

 **Team Leader**
Data Science Practice

 **Data Engineers**
Data Scientists
“Code First”
/ Package First

Open for Innovation
 **KNIME**

 **Business Analyst**
Interactive User
Information Consumer

 **Model / ML Operations**

 **IT Operations**
Scalability and control
Centralized resources / strategies
Security

 **Compliance**
Financial / risk oversight
Costs allocation
Governance, traceability, GDPR

THE END
