Welcome to
Sharing and Deploying Data Science with KNIME Server

Going live at:

Berlin 5:00 PM (CEST)
New York City 11:00 AM (EDT)
Austin 10:00 AM (CDT)
London 4:00 PM (GMT)
Before we start...

- Please use the Q&A section to post your questions.
- Upvote for your favourite questions.
- Session is recorded and will be available on YouTube.
KNIME Server

• Collaboration
  – Share Expertise and Templates/Blueprints

• Automation
  – Schedule, Monitor, Update

• Deployment
  – To Humans: Analytical Applications ("Guided Analytics")
  – To Machines: via REST and custom APIs

• Management
  – Manage distributed setups
  – Integrate with existing enterprise setup (authentication)
Challenges Operationalizing Data Science

Creating Great Data Science

- Large, complex variety of technologies
- Providing skills and resources

Data Science Driven Applications

- R&D
- MFG
- IT
- Marketing
- Sales
- ... more

Continuous, high quality delivery
Mainstreaming across the organization
Challenges Operationalizing Data Science

Data Science Driven Applications

- R&D
- MFG
- IT
- Marketing
- Sales
- ... more

Productionizing Data Science

- Continuous, high quality delivery
- Mainstreaming across the organization

Large, complex variety of technologies
Providing skills and resources
Better Decision-making, Faster!

Load > Integrate > Transform > Analyze > Visualize

KNIME Analytics Platform

KNIME Integrations

KNIME Extensions

Community Extensions

Partner Extensions

KNIME Server

Collaborate — Automate — Manage — Deploy
KNIME Server

Collaboration

Automation

Deployment

Management

Workflow Hub

Workflows

Metanodes, Nodes

Data

Shared Repository, Versioning

Provisioning

Remote, Client Control

Distributed Executors

Web Front-ends

Guided Analytic Apps

RESTful Services

Web Browser Access

Custom Application Access

Security

Configuration & Versioning

Client & User Preferences

© 2020 KNIME AG. All rights reserved.
Live Demo
Guided Analytics for ML/AI Automation

- Interaction & Automation
- Data Scientist’s Choice: right mix for target audience
Compare Models Time

Compare in those bar charts the time each model took to train and to compute a sample prediction.

Training Time

- Decision Tree
- Logistic Regression
- Naive Bayes
- Random Forest
- Gradient Boosted Trees
- Deep Learning

Prediction Time per Sample

- Naive Bayes
- Deep Learning
- Random Forest
- Decision Tree
- Gradient Boosted Trees
- Logistic Regression

Details for Experts

You can compare the models now by different metrics. Above you can see a bar plot with 3 main measures of performance: Accuracy is the percentage of correct predictions among all predictions. Area Under Curve (AUC) measures of the area under the Receiver Operating Characteristic curve.

The ROC plot describes the Receiver operating characteristic curves, one for each model. On the y-axis you have the true positive rate, on the x-axis you have the false positive rate based on the class you previously provided. The greater the area under the model curve (AUC) the better the performance of the model. More info about this plot can be found at this link.

Advanced Assessment of Models

The advanced assessment of models sections shows four additional charts per model:

- **1. Performance Metrics Bar Charts**
  
  For this visualization we measured the following metrics:
  
  - **Recall** (or True Positive Rate) (% of "casualty" rows correctly classified)
  - **Precision** (or Positive Predicted Value) (% of predicted "casualty" rows correctly classified)
  - **Specificity** (or True Negative Rate) (% of not "casualty" rows correctly classified)
  - **F-norm** (harmonic average between Recall and Precision)

- **2. Cumulative Gain Chart and Lift Chart**
  
  This chart is drawing a curve that reflects how well the model is doing compared to a random classifier. You are selecting rows from the test ranked by the probability of class "casualty". On the x-axis you have the percentage of top ranked rows by the model that define the partition of rows you are considering. On the y-axis you measure the response as the percentage of "casualty" rows over their total number in your partition of top ranked rows. If the model is bad, the curve will be close to the black line, where the percentage of original "casualty" rows is exactly equal the percentage of selected rows (e.g. where the model ranking is random). The cumulative gain curve should be above the bisector line and the greater the area between the cumulative gain curve and the bisector line is, the better the model is.
Advanced Assessment of Models

Each row represents a series of additional information about each created model.

Decision Tree

Figure 1: Performance Metrics
- Recall (TPR)
- Specificity (TNR)
- Precision (PPV)
- F-measure (F1)

Figure 2: Cumulative Gain Chart

- Cumulative Gain
- random

Figure 3: Global Feature Importance
- Pclass
- Sex
- Embarked
- Age
- SibSp
- Parch
- Fare
- row ID
- Cabin
- Ticket
- Name

Figure 4: Confusion Matrix Heatmap

Gradient Boosted Trees

Naive Bayes
ROC Curve Plots

Compare the curves, one for each model. The greater the area under a curve the better the model is.

![ROC Curve Graph]

- **Decision Tree (0.823)**
- **Gradient Boosted Trees (0.582)**
- **Naive Bayes (0.85)**
- **Random Forest (0.861)**
- **Logistic Regression (0.510)**
- **Deep Learning (0.789)**

Download Model

The following table summarizes the information in the charts. Please select from the table the model you would like to download and use.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%)</th>
<th>Training Time (mins)</th>
<th>Prediction Time (millsec)</th>
<th>Area Under Curve (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>94.276</td>
<td>0.4</td>
<td>3</td>
<td>82.323</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>86.819</td>
<td>0.4</td>
<td>3.4</td>
<td>51.515</td>
</tr>
<tr>
<td>Random Forest</td>
<td>84.812</td>
<td>0.8</td>
<td>2.9</td>
<td>88.062</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>82.713</td>
<td>0.4</td>
<td>2</td>
<td>85.047</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>81.28</td>
<td>4</td>
<td>2.6</td>
<td>78.881</td>
</tr>
<tr>
<td>Gradient Boosted Trees</td>
<td>60.036</td>
<td>0.9</td>
<td>3.1</td>
<td>58.226</td>
</tr>
</tbody>
</table>
KNIME Server REST API

- Enables external integration
- Build applications around KNIME Server
- e.g. for Microservices and real time scoring

Deploy KNIME workflows as web services

```
"Day Mins":191.5,
"Eve Mins":175.2,
"Night Mins":220.3,
"Int1 Mins":0.0
```
KNIME Server REST API: Job Pool

KNIME Workflow for e.g. sentiment prediction, churn score, chemical property

External REST call

REST API

KNIME SERVER

Take from pool

Pooled jobs

KNIME Executor

Workflow immediately executed

Pre-loaded jobs pooled and waiting for work

Results

No “load latency” for client

© 2020 KNIME AG. All rights reserved. 24
Managing Preferences – local

How do I configure Python?

How do I install database drivers?

KNIME Website
Internal knowledgebase
Managing Preferences – KNIME Server

- Different departments/teams have different requirements
- Multiple OS deployments
  - Windows 7
  - Windows 10
  - Linux
  - macOS

Marketing

- Windows 10
- Hive
- Spark

Finance

- Windows 7
- Oracle
- MS Access

R&D

- Linux and macOS
- Python
- R
Managing Preferences – KNIME Server

Profiles can include:
- Client-profiles
  - Python-Linux
  - Python-macOS
  - R-Linux
  - R-macOS
  - Databases-Win7
  - Big Data-Win10

Profiles can include:
Preferences, drivers, and more
KNIME Server – Admin made easy

![KNIME Administration interface](image)

### Status Info

<table>
<thead>
<tr>
<th>Host: LAPTOP-87K240HL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uptime: 0 days, 0 hours, 1 minutes, 59 seconds,</td>
</tr>
<tr>
<td>Executors Info: KNIME Executors associated with user: (none): 0: RMI Executor 0 (user: (none), port: 50100, uptime: 00:00:52, status: AcceptingJobs, #jobs: 0)</td>
</tr>
</tbody>
</table>

### Download Server Log Files

### License and Users Info

- **License Type: KNIME Server Large**
- **Expiration Date: 2019-12-31**
- **Company: KNIME**
- **Customer: Roland Burger**
- **Maximum Cores: 8**
- **Host Identifiers:**
  - "Host MAC": "F0:D5:BF:EF:3D:27"
  - "Computer": "Server Large Test License"

### Accounts

<table>
<thead>
<tr>
<th>State</th>
<th>In Memory</th>
<th>Name</th>
<th>Workflow</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTED</td>
<td>no</td>
<td>KNIME_project 2018-08-01 11.37.48</td>
<td>KNIME_project</td>
<td>knimeadmin</td>
</tr>
<tr>
<td>EXECUTED</td>
<td>no</td>
<td>01_CustomerSegmentation_ModelLearning 2018-07-31 15.25.30</td>
<td>/00_ServerWebinar/03_workflows/01_CustomerSegmentation_ModeILearning</td>
<td>knimeadmin</td>
</tr>
<tr>
<td>CONFIGURED</td>
<td>no</td>
<td>01_CustomerSegmentation_UseCase 2018-07-28 11.37.04</td>
<td>/00_ServerWorkspace/00_Demo/03_CustomerSegmentation_UseCase</td>
<td>knimeadmin</td>
</tr>
<tr>
<td>EXECUTED</td>
<td>no</td>
<td>KNIME_project 2018-08-01 11.38.42</td>
<td>KNIME_project</td>
<td>knimeadmin</td>
</tr>
<tr>
<td>EXECUTED</td>
<td>no</td>
<td>01_CustomerSegmentation_ModelLearning 2018-08-01 14.42.04</td>
<td>/00_ServerWebinar/03_workflows/01_CustomerSegmentation_ModeILearning</td>
<td>knimeadmin</td>
</tr>
</tbody>
</table>
Distributed Executors

• Need more workflow horsepower?

• KNIME Server currently supports ‘Scale Up’

• KNIME Server Distributed Executors allows ‘Scale Out’
Distributed Executors

Requests to execute workflows

KNIME Server

RabbitMQ

Executor 1
Executor 2
Executor 3

KNIME Server + Distributed Executors
Distributed Executors

Request to execute workflow

KNIME Server

RabbitMQ

Scripted launch on AWS CloudFormation

Executor 1

KNIME Server + Distributed Executors
Distributed Executors

More requests to execute workflows

KNIME Server

RabbitMQ

Executor 1
CPU load increases

Launching Executor 2

AWS Auto-scaling Group

KNIME Server + Distributed Executors
Distributed Executors

Requests to execute workflows

 KNIME Server

 RabbitMQ

 Executor 1

 Executor 2

 AWS Auto-scaling Group

 KNIME Server + Distributed Executors
Workflow Pinning

• Example Scenarios:
  – My workflow needs access to the Finance database
  – My workflow requires GPU’s and Python

• Pinning Workflows
  – Tag a deployed workflow
  – Execute on Server
  – Workflow routed to matching Executor
KNIME Software – On premise and in the cloud
KNIME Server

Collaboration
- Workflow Hub
- Workflows
- Metanodes, Nodes
- Data
- Shared Repository, Versioning

Automation
- Provisioning
- Remote, Client Control
- Distributed Executors
- Scheduled Execution

Deployment
- Web Front-ends
- Guided Analytic Apps
- RESTful Services
- Web Browser Access
- Custom Application Access

Management
- Nodes, Files, Applications
- Security
- Configuration & Versioning
- Client & User Preferences

Ready to get started with KNIME Server? Contact us for a free trial:

contact@knime.com
The End –
thank you for joining this webinar.