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
Introduction to KNIME Analytics Platform and KNIME Server

Paolo Tamagnini, KNIME
KNIME® Software

Load ➔ Integrate ➔ Transform ➔ Analyze ➔ Visualize

Open Source

KNIME Integrations

KNIME Analytics Platform

Community Extensions

Partner Extensions

KNIME Server

Collaborate ➔ Automate ➔ Manage ➔ Deploy

© 2019 KNIME AG. All rights reserved.
Guided Analytics Learnathon
Building Applications for Automated ML

Paolo Tamagnini, KNIME
The CRISP-DM Cycle

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment
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?
The CRISP-DM Cycle with Wrapped Metanodes

- Load Data
- Clean & Transform Data
- Analyse
- Visualize
- Deploy
Examples of Wrapped Metanode Views

1st Period Time Range

- Use Only 1st Period
- From: 2017-01-01
- To: 2017-12-31

Upload Dataset

Upload the dataset to be used.

Uploading file "adult.csv"

Download the Labeled Data

Download the labeled .csv file with labeled data.

Load Data → Clean & Transform Data → Analyse → Visualize → Deploy
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
      - 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)
    - 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

Using Interactivity of JavaScript nodes in a Composite View!
Customizing the View Layout
Connecting to KNIME Server

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 Workflow to KNIME Server

Choose the destination workflow group:

- knime-server-learnathon (paolo.tam@https://learnathon.knime.com/knime/rest)
  - 01_Guided_Analytics_for_ML_Automation
  - Admin
  - admin_live_demo
  - deploy
  - Examples
  - other
    - user-0
    - user-1
    - user-10
    - user-11
    - user-12
    - user-13
    - user-14
    - user-15
    - user-16
    - user-17
    - user-18
    - user-19
    - user-2
    - user-20
    - user-21
    - user-22

- knime://knime-server-learnathon/user-0

Reset Workflow(s) before upload

[OK] [Cancel]
Composite View from KNIME WebPortal

From the workflow deployed on KNIME Server..

Composite View from KNIME WebPortal

- Bar Chart
- Parallel Coordinates Plot
Interactivity across Charts: Selection and Filter Events

WrappedNode Input

Parallel Coordinates Plot

Bar Chart

Publishing Selection

WrappedNode Output
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

1. **Data Files**
   - Database

2. **Data Blending**
   - Cloud

3. **Human Interactions**

4. **Machine Learning Automation**

5. **Download Models**
Machine Learning Automation

Selected Data

Data Preparation → Feature Engineering

Features Selection → Parameter Optimization

Model Retraining → Model Evaluation

Model Tuning

Download Models
Today Exercise: a simple example of ML automation
Today Exercise: Adult Dataset

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

- Age
- Workclass
- fnlwgt
- Education
- Education-Num
- Marital Status
- Occupation
- Relationship
- Race
- Sex
- Capital Gain
- Capital Loss
- Hours per week
- Country
- **Income**
- Gender with Income

$> 50K$

$\leq 50K$
Exercise 1: User Interface for Automated ML Settings

Group 1 Challenge
Exercise 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]
Exercise 2: Dashboard with Models Performance

Group 2 Challenge
Exercise 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!

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..

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

![KNIME Data Generation](image1)

<table>
<thead>
<tr>
<th>Extension</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNIME Data Generation</td>
<td>3.7.0.v201808081048</td>
</tr>
<tr>
<td>KNIME H2O Machine Learning Integration</td>
<td>3.7.1.v201901281201</td>
</tr>
<tr>
<td>KNIME JavaScript Views (Labs)</td>
<td>3.7.1.v201901281201</td>
</tr>
<tr>
<td>KNIME PMML Preprocessing Applier Nodes</td>
<td>3.7.0.v201808081048</td>
</tr>
</tbody>
</table>
Open Exercise 1: Automation Setup
Open Exercise 2: Automation Results
Ideally...

Data Scientist

Business Analysts

That’s it! Perfect!

Here we go!
User Interactions

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

Upload Data and Process Setup
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

Fine-tune Model Parameters or Use Automatic Settings
If you selected to fine-tune the model parameters:
1. Define the parameter values used for optimization
2. Define feature engineering settings
   Otherwise the settings will be automatically set.

Execution Settings
Decide which execution environment you want to use.

ML Automation
Model parameters are automatically optimized and features engineered.

Download Models
Compare and inspect the results of the models and download the desired ones.
A Blueprint for Guided Automation
The Workflow

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

2. Select Target
   - Select the target column whose value is to be predicted.
   - Select: Label

3. Filter Columns
   - Set Column Relevance Filter
   - Use the slider to select a subset of columns.
   - Filtered value: 4.55

4. Feature Quality Calculation
   - Calculate the quality of features.

5. Select Models
   - Choose one or more machine learning models.
   - Simple models:
     - Naive Bayes
     - Decision Tree
     - Logistic Regression
The Workflow

Random Forest
Set the parameter ranges for the Random Forest.

Number of Trees

Maximal Depth

Feature Engineering Settings
Select Techniques
Please select the feature engineering techniques you want to use.
- Simple Transformations
- Feature Combinations
- Dimensionality Reduction

Execution Settings
Available options:
- Local execution
- Use Spark cluster if possible
- Use Apache Spark MLlib
- Use other cluster environment
The Workflow

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 visualizes different performance metrics to assess the quality of each model.
On the Example Server..

EXAMPLES/50_Applications/36_Guided_Analytics_for_ML_Automation/01_Guided_Analytics_for_ML_Automation
On the Workflow Hub..

Guided Analytics for Machine Learning Automation

This workflow generates 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 will generate a web page with which the business analyst can interact.
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 YouTube
• KNIME on @KNIME
• KNIME on https://www.facebook.com/KNIMEanalytics
• On Meetup
Thank You!

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.