[L2-DW] KNIME Analytics Platform for Data Wranglers: Advanced

April 17, 2023

Structure of the course

- This course consists of five sessions
 - Data/Time and Data Export
 - Flow Variables and Shared Components
 - Workflow Control: Loops, Switch, and Try-Catch
 - Formatting Excel Tables and Introduction to Data Science
 - Q&A and Summary
- Structure of each session
 - Introduction to the topic
 - Hands-on exercise
 - Solution walk-through



Downloading the Exercises

 Download the course material from the KNIME Community Hub https://hub.knime.com/hayasaka/spaces/KNIME%20Spring%20Summit%20Training%202023/latest/





Importing the Exercises

Import the course material to KNIME Analytics Platform



4

Date/Time Data



Date & Time Overview

- Dedicated data type for date and time data
- Supported in Date&Time nodes
 (and others: GroupBy, Pivot, Line Plot)
- Complete re-write in KNIME 3.4



Other Data Types Setwork Text Processing Time Series Manipulate Create Date&Time Range Date&Time Difference Date&Time Shift Date&Time-based Row Filter Modify Date Modify Time Modify Time Zone Transform ☑ Date&Time to String S String to Date&Time るUNIX Timestamp to Date&Time Contraction to String 🖓 Duration to Number Sh String to Duration 2 Date&Time to legacy Date&Time Legacy Date&Time to Date&Time Extract Date&Time Fields Extract Duration Fields 🚍 Window Loop Start Smoothing Moving Aggregation Moving Average



String to Date&Time

- Converts date/time da into a native Date&Tin
- Guesses correct form types of date formattir
 - Enter format manually if a didn't work
 - KNIME automatically formats to auto-guess
 - Converts multiple columns format in one node

ate/time data from string	Dialog - 0:241 - String to Date&Time (Transform Transform Transfor
e formatting manually if auto-guessing	Manual Selection Wildcard/Regex Selection Include Filter S City S CustomerID S FirstName S Birthday S Email S Newsletter ProductNr
automatically adds custom to auto-guess list Itiple columns of same date	S Type Enforce exclusion Replace/Append Selection Append Selected columns Suffix of appended columns: (Date&Time) Banually
e node	Type and Format Selection New type: Date Date format: M-d-yyyy Locale: n-GB Content of the first cell: 11-2-2014
String to Date&Time	Guess data type and format
Select type of output column	Abort Execution Click to auto-guess format
	OK - Execute Apply Cancel



Date&Time – Data Types

e Hilite		
	Navigation View	
Ta	able "default" – Rows: 1000	۲
Row ID	Date	
Row0	2017-01-19	0
Row1	2017-01-19	
Row2	2017-01-20	
Row3	2017-01-20	
Row4	2017-01-20	
Rows	2017-01-21	
Row7	2017-01-22	
Row8	2017-01-22	
Row9	2017-01-22	
Row10	2017-01-23	
Out	nut table - 2:50 - Create Date&Time Bang	
😑 🔵 Out le Hilite	put table - 2:50 - Create Date&Time Rang Navigation View	le
● ● Out e Hilite	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000	ie
Out Ie Hilite Tr Row ID	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000	le
Out E Hilite Row ID Row0	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 70 Date&Time 2017-01-19T13:00:46	
Out Hilite Row ID Row1	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 70 Date&Time 2017-01-19T13:00:46 2017-01-19T21:46:57	je
Out le Hilite Row ID Row0 Row1 Row2	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 [70] Date&Time 2017-01-19T13:00:46 2017-01-19T21:46:57 2017-01-20T06:33:08 2017-01-20T06:33:08	le
Out Ie Hilite Row ID Row0 Row1 Row2 Row3 Pow4	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 [7] Date&Time 2017-01-19T13:00:46 2017-01-20T06:33:08 2017-01-20T06:33:08 2017-01-20T06:31:09:20	je
Out E Hilite Row ID Row0 Row1 Row2 Row3 Row4 Row4	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 70 Date&Time 2017-01-19T13:00:46 2017-01-19T21:46:57 2017-01-20T06:33:08 2017-01-20T05:19:20 2017-01-21T08:51:42	le le
Out Hilite Row1 Row2 Row2 Row3 Row4 Row5	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 Date&Time 2017-01-19T13:00:46 2017-01-19T21:46:57 2017-01-20T156:57 2017-01-20T15:19:20 2017-01-21T0:51:14 2017-01-21T17:53	le
Out Hilite From ID Row1 Row2 Row3 Row4 Row5 Row6 Row7	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 70 Date&Time 2017-01-19T13:00:46 2017-01-20T06:33:08 2017-01-20T06:33:08 2017-01-21T00:05:31 2017-01-21T08:51:42 2017-01-21T07:53 2017-01-22T02:24:04	
Out Hilite Tr Row ID Row2 Row3 Row5 Row6 Row7 Row8	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 [7] Date&Time 2017-01-19T13:00:46 2017-01-20T06:33:08 2017-01-20T06:33:08 2017-01-20T06:531 2017-01-21T08:51:42 2017-01-21T08:51:42 2017-01-22T02:24:04 2017-01-22T11:10:15	
Out Ie Hilite Row1 Row2 Row3 Row4 Row5 Row6 Row7 Row8 Row9	put table - 2:50 - Create Date&Time Rang Navigation View able "default" - Rows: 1000 70 Date&Time 2017-01-19T13:00:46 2017-01-19T21:46:57 2017-01-20T06:33:08 2017-01-20T05:31 2017-01-21T08:51:42 2017-01-21T08:51:42 2017-01-22T02:24:04 2017-01-22T112:015 2017-01-22T19:56:27	

	Flow Variables Job Manager Selection Memory Policy
Output Settings	s Date
Output type:	✓ Date&time
	Date&time with zone
New column na	ame: Date&Time
Mode Selection	1
Number of row	vs: D Fixed: 1 000 ^
	Variable
Starting Point	
Start:	Date: 2017-01-19 Time: 13:00:46
	Time Zone: Europe/Berlin 🗘
	Use execution datedtime
Ending Point	
Ending Point — O Interval:	
Ending Point — Interval:	
Ending Point Interval: End:	Date: 2018-01-19 Time: 14:00:46
Ending Point	Date: 2018-01-19 Time: 14:00:46



Row10



2018-01-19T14:02:31.155+01:00[Europe/Berlin]

Date & Time + Time zone

String to Duration

- Takes a string and converts it to a duration cell
 - Three different options to format input strings
- Example: Convert 1 year, 2 months, 3 weeks, and 4 days to duration cell
 - ISO-8601: "P1Y2M3W4D"
 - Short letter: "1y 2M 3w 4d"
 - Long word: "1 year 2 months 3 weeks 4 days"

Exclude	Selection Wildcard/Regex Selection
T Filter	Tilter
No columns in this list	S Date S Time S timestamp
Enforce exclusion	• Enforce inclusion
 Append selected columns Suffix of Replace selected columns 	of appended columns: (Duration)
Type Selection	
 Automatically detect type 	
Automatically detect type Create date-based duration	
 Automatically detect type Create date-based duration Create time-based duration 	



			Outpu	t table - 0:51 - String	to Duration		
File Hilite	Navigation	View					
		Table "default"	- Rows: 1	Spec – Columns: 6	Properties	Flow Variables	
Row ID	S iso	S short	S long		iso(Duration)	short(Duration)	long(Duration)
Row0	P1Y2M3W4	D 1y 2M 3w 4d	l 1 year 2 m	onths 3 weeks 4 days	1y 2M 25d	1y 2M 25d	1y 2M 25d



Date&Time-based Row Filter

- Filters rows from a specified time period
- Range can be limited on upper bound, lower bound or both
- Options for end point:
 - Date&Time: Fixed data and time
 - Duration: Duration string (e.g. 2y 3M)
 - Numerical: Select granularity from dropdown and enter number

ate&Time Se	lection
☑ Start:	Date: 2007-01-01 Image: Time: 00:00:00 ▲ ▲ Now Time Zone: Europe/Berlin ♦ ✓ Inclusive Use execution date&ime
🗹 End:	● Date&Time Date: 2007-01-31 ■ Time: 23:59:59 → ■ Now ● Duration Numerical Time Zone: Europe/Berlin ● ■







Extract Date&Time Fields

- Extracts date fields (year, day, month) or time fields (hour, minute, second) from a date&time cell.
- You can pick and choose which fields to include
- Useful when used in combination with data aggregation nodes (groupby, pivot etc.)



Date&Time column 🗊 Date 🗘	
Date Fields Year Year Year Quarter Output Month (number) Month (name) Week Day of year Day of month Day of week (number) Day of week (number)	Time Fields Hour Ninute Second Subsecond in milliseconds C Time Zone Fields Time zone name Time zone offset
Localization (month and day names, e Locale en-GB	tc.)



Date&Time Difference

- Check the difference between a time column and...
 - Another time column
 - Execution time
 - User-defined time
 - Time from previous row
- Choose desired resolution (days, hours, minutes, etc.)

Date&Time Difference

To cal colu

	Options	Flow Variables	Job Manager Selection	Memory Policy
-Base colun	าท			
Date&Tim	e column	🖥 Time ᅌ		
-Calculate d	lifference to			
o seco	nd column		🔁 Tin	ne 📀
🔵 curre	nt execution d	ate&time		
fixed	date&time			
O previ	ous row			
Output opt	ions			
O Gran	larity	Davs		
Durat	ion			
	-			
New colu	imn name: d	ate&time diff		
			OK Appl	y Cancel 🕜
		e ve el		
amerence	e to sec	ona		
th column	s need	to		

Date&Time Shift

- Shifts date or time by either a duration or a numerical value
- Use duration:
 - Use duration column
 - Or shift by user-defined value
 - E.g. 1y, 2M, 5h, etc.
- Use numerical in combination with user-defined granularity

Date&Time Shift

- Use numerical column
- Or shift by user-defined value

Exclude	
Filter	T Filter
No columns in this list	> I timestamp
O Enforce exclusion	C Enforce inclusion
Append selected columns Suffix of appe Replace selected columns Shift Value Selection Use Duration	nded columns: (shifted)
	^
Duration column Duration value	v)
Duration column Duration value Use Numerical	v
Duration column Duration value Use Numerical Numerical column	\$
Duration column Duration value Use Numerical Numerical column Numerical value 1	¢ •

Modify Time / Modify Date

- Modify Date&Time columns
- Three options:
 - Appends time (date) to date (time) column
 - Changes time (date) to fixed value
 - Removes time (date) from Date&Time column
- Column selection shows only columns suitable for currently selected option



Exclude	💽 Manual Se	election 💛 Wildca	ard/Regex Selection	
T Filter			TFilter	
No colun	nns in this list		T timestamp	_
		»		
		«		
			C Enfance inclusion	
eplace/Append Select	ion			
Append selected c	olumns Suffix of	appended column	s: (modified date)	
Replace selected c	olumns			
ate Selection				
Append date	Date: 2018 (02 🗖		
Change date	Date. 2018-0	J8-02		
-	✓ Time Zone	: Europe/Berli	n 🗘	
Remove date				

Modify Time Zone

- Similar to Modify Time/Modify Date
- Input: Date&Time
 - Set time zone
- Input: Date&Time (Time zone)
 - Set time zone

© 2023 KNIME AG. All rights reserved.

- Shift time zone
- Remove time zone

Dialc	og - 0:51 - Modify	Time Zone
Options	Flow Variables	Memory Policy
O Manual Se	lection 🔵 Wildc	ard/Regex Selection
T Filter		T Filter
No columns in this list	$\mathbf{>}$	😰 timestamp
	»	
	<u>«</u>	
Enforce exclusion		O Enforce inclusion
Append selected columns Suffix of a Replace selected columns ime Zone Selection Set time zone Time Zone:	Europe/Berlin	s: (modified time zone)
Shift time zone Remove time zone		
	c	DK Ap Cancel 🝞

Modify Time Zone







Exercise: 08_DateTime_Manipulation

- Convert order dates from string to Date&Time
- Extract the product purchases that were submitted in 2019
- Extract the remaining product purchases into a separate table
- Extract monthly and year of each product purchase into separate columns
- Plot monthly sales in a line plot



Exporting Data & Deployment



Exporting Data

After an analysis is completed, what next?

- Write results to a file
- Upload results to a Cloud Storage
- Create/update a database
- Generate a rich report using BIRT
- Send your data to Tableau, Spotfire, PowerBI to create a report
- Deploy via KNIME KNIME Business Hub
- Deploy your model as RESTful web service



Data Export Nodes

- Typically characterized by:
 - Magenta color
 - 1 input port, no output ports
 - Create file on file system or write to database







V <> Structured Data VOSC NOSC JSON Writer JSON Row Combiner and Writer ▼ set XML * XML Writer ▼ (···) Scripting V Python Python Object Writer R 🔥 R V . 10 R Model Writer V 😵 Community Nodes V S KNIME Image Processing V 🏫 10 -Image Writer V Erlwood Nodes V 10 XLS Coloured Writer V III NGS V 10 A BEDGraphWriter FQCWriter FastQWriter 🔻 💀 Palladian Text Classifier TextClassifierModelWriter T RDKit **Fingerprints** RDKit Fingerprint Writer

V 📩 IO Vrite CSV Writer **ARFF ARFF Writer** . I Table Writer PHIL PMML Writer A Model Writer + Image Port Writer + Image Column Writer *LS Excel Writer (XLS) Analytics V Mining 🔻 🐗 Weka Veka (3.7) V 🔾 IO Weka Classifier Writer (3.7) Weka Clustering Writer (3.7) ▼ ↔ Distance Calculation V 📒 Distance Matrix ∃→ Distance Matrix Writer V S Database Read/Write Database Connection Table Writer ← Database Writer V Other Data Types Vetwork V 10 ▼ 🗐 File Solution Network Writer

https://www.youtube.com/watch?v=Og7VZOJhsOc&feature=youtu.be



Table Writer



		Setting	s Flow Variables			
Write to	Relative to	Current wo	orkflow ᅌ			
File	//data/temp/Pre	ediction.table			S Browse.	
Write options	Create missing f	folders If exists:	overwrite 🔵 fail	Ê.		
			01/	Analy	Canad	



Excel Writer

- Writes the input table into a spreadsheet of an Excel file
- Select append, to append a spreadsheet to an existing Excel File and define the name of the new sheet



	Dialog - 4:69 - Excel W	/riter (Write records)
S	ettings Flow Variables	Job Manager Selection
File format & output location		
Excel format XLSX ᅌ		
Write to Relative to	Current w	vorkflow
File//data/tem	p/products.xlsx	Srowse
Write options Create miss	ing folders If exists	5: 🔵 overwrite 💿 append 🔵 fail
<u>cl</u>		
Sneets		
1. sneet name derault_1		
If sheet exists 📀 overwrite	🔵 fail	Excel Sheet
Names and IDs		appender
write row key		
Write column headers		
Missing value handling		
Replace missing values by		
Formulas		
Evaluato formulas		
Evaluate formulas		
Layout		
Autosize columns		
○ Portrait ○ Landscape	A4 – 210x297 mm	≎
	OK - Execu	ite Apply Cancel 🕐



Write Files to a Remote File System

- The new file handling framework makes it easy to upload data to remote file systems
 - Write processed data directly with a writer node
 - Upload local files with the Transfer Files node
- Supported file systems
 - Microsoft Azure
 - Google
 - Amazon
 - Databricks
 - BigData file systems (hdfs, httpFS, ...)
 - On-premise (e.g. ssh, ftp, ...)





CSV Writer

Full Flexibility with the Transfer Files node



Same cloud environment





On-premise



Cross cloud environments



Other Utility Nodes

Can be used with local and remote file systems

- Create a folder
- Delete files or folders
- List all files in a folder



Further information about file handling

https://docs.knime.com/latest/analytics_platform_file_handling_guide/index.html





DB Writer

 Writes data from a KNIME data table directly into a database table File Reader

Settings Output Type Mapping Flow Variables	Job Manager Selection M	lemory Policy		_
Table to write				
Schema:	Table: Cur	rentDetailData	Select a table	
Batch Size: 1,000	Fail on error	Append write status columns	Remove existing table	
Select the columns to write (Tin SQL)				
Exclude	Selection () Wildcard/Reg	ex Selection Type Selection Type Selection		
Titer		Teiter		
		S MaritalStatus		
rease batch size for	/	S Gender	Append to	o or d
better performance	>>	I NumberOfContracts	existing	g table
	<	I Age S Target		
		Available401K		
	~~	D ChurnScore	~	
Enforce exclusion		Enforce inclusion		

tapping by Name Column Name Regex Source Type Mapping tapping by Type tring → String → VARCHAR oplan value → SolcEAN	tapping by Name Column Name Regex Source Type Mapping	ttings Output	t Type Mapping	Flow Variables	Job N	lanager Selection	mory Policy		
Column Name Regex Source Type Mapping 🗭	Column Name Regex Source Type Mapping Iapping by Type OUME Type OUME Type Solean value oolean value mber (integer) Integer → INTEGER	Mapping by Nar	me						
Mapping by Type KNUME Type	Iapping by Type Image: String → VARCHAR OUME Type Mapping Outer Tring → String → VARCHAR Solean value → Boolean → BOOLEAN umber (integer) → Integer → INTEGER umber (integer) → Long → BIGINT	Column Name		R	legex	Source Type	Mapping	0	
Mapping by Type KNIIME Type Mapping String → String → VARCHAR Soolean value → Boolean → BOOLEAN	Iapping by Type Mapping ● NUME Type Mapping ● tring → String → VARCHAR ▲ oblean value → Boolean → BOOLEAN ▲ umber (integer) → Integer → INTEGER ↓ unber (ong) → Long → BIGINT ▲								^
Mapping by Type KNIME Type Mapping ♥ Itring → String → VARCHAR Sociean value → BOolEan → BOOLEAN	Iapping by Type VIME Type Mapping ● NUME Type Mapping ● ● tring → String → VARCHAR ▲ obclean value → Boolean → BOOLEAN ▲ umber (integer) → Integer → INITEGER ▲ unber (ong) → Long → BIGINT ▲								
Aapping by Type NNUME Type Mapping tring → String → VARCHAR oolean value → Boolean → BOOLEAN	Implicit Section Implicit Section Mapping by Type DNIME Type Mapping Implicit Section Implici								
Mapping by Type KNUME Type Mapping Image: Comparison of the system of the sys	tapping by Type XNME Type Mapping tring → String → VARCHAR oblean value → Boolean → BOOLEAN umber (integer) → Linteger → INTEGER under (ong) → Long → BIGINT								
Mapping by Type Mapping Image: Constraint of the second	tapping by Type NUME Type Mapping tring → String → VARCHAR polean value → Boolean → BOOLEAN umber (integer) → Integer → INTEGER umber (integer) → Long → BISINT								
KNIME Type Mapping Image: Constraint of the state of	Mapping Mapping NUME Type Mapping Solean value Solean -> BOOLEAN umber (integer) -> Integer -> INTEGER umber (integer) -> Long -> BIGINT								~
String → String → VARCHAR	→ String → VARCHAR ∧ polean value → Boolean → BOOLEAN umber (integer) → Integer → INTEGER umber (integer) → Long → BISINT	Mapping by Typ)e						~
Boolean value → Boolean → BOOLEAN	oolean value → Boolean → BOOLEAN umber (integer) → Integer → INTEGER umber (integer) → Long → BIGINT	Mapping by Typ KNIME Type	be			Мар		0	~
	umber (integer) → Integer → INTEGER umber (integer) → Long → BIGINT	Mapping by Typ KNIME Type String	be			Map → St	→ VARCHAR	0	~
Number (integer) → Integer → INTEGER	umber (long) → Long → BIGINT	Mapping by Typ KNIME Type String Boolean value	De			Mapi → Str → Bo	→ VARCHAR → BOQLEAN	0	^
Number (long) \rightarrow Long \rightarrow BIGINT		Mapping by Typ KNIME Type String Boolean value Number (integer	r)			Mapı → Str → Bo → Ini	→ VARCHAR an → BOOLEAN r → INTEGER	¢	^
Number (double) → Double → DOUBLE	umber (double) → Double → DOUBLE	Mapping by Typ KNIME Type String 3oolean value Number (integer Number (long)	r)			Maple \rightarrow St \rightarrow Bo \rightarrow In \rightarrow Lo	→ VARCHAR an → BOOLEAN ar → INTEGER → BIGINT	0	^
		Mapping by Typ KNIME Type String Boolean value Number (integer Number (long) Number (double	r)			Mapı → Stı → Bo → In → Lo → Do	→ VARCHAR an → BOOLEAN er → INTEGER → BIGINT = → DOUBLE	0	^
\rightarrow LOCAIDALE \rightarrow DATE	ocal Date → LocalDate → DATE	Mapping by Typ KNIME Type String Boolean value Number (integer Number (long) Number (double Local Date	r) 2)			$\begin{array}{c} Map \\ \rightarrow St \\ \rightarrow Bo \\ \rightarrow In \\ \rightarrow Lo \\ \rightarrow Da \\ \rightarrow Lo \end{array}$	→ VARCHAR an → BOOLEAN er → INTEGER → BIGINT e → DOUBLE abt → DATE	0	
Local Time → LocalTime → TIME	ccal Date → LocalDate → DATE	Mapping by Typ KNIME Type String Boolean value Number (integen Number (iong) Number (double ocal Date ocal Time	r)			Mapı → St → Bo → In → Lo → Lo → Lo	→ VARCHAR an → BOOLEAN ar → INTEGER → BIGINT = → DOUBLE bate → DATE ime → TIME	0	
tocal Jose → UocalJose → UA1E Local Time → LocalTime → TIME Local Date Time → LocalDateTime → TIMESTAMP	ccal Date → LocaDate → DATE scal Time → LocaTime → TIME scal Date Time → LocaDateTime → TIMESTAMP	Mapping by Typ KNIME Type String Boolean value Number (integei Number (double .ocal Date .ocal Time .ocal Time	r) 2)			Mapı → Stı → Bo → In' → Lo → Lo → Lo → Lo → Lo → Lo	→ VARCHAR → VARCHAR an → BOQLEAN ar → INTEGER → BIGINT = → DOUBLE bate → DATE ime → TIME takeTime → TIME takeTime → TIMESTAMP		*
Number (integer) → Integer → INTEGER. Number (long) → Long → BIGINT Number (double) → Double → DOUBLE	umber (double) → Double → DOUBLE	Mapping by Typ KNIME Type	De			Мар		0	
acal Data		Mapping by Typ KNIME Type String Soolean value Number (integer Number (long) Number (double	r)			$\begin{array}{c} Mapi \\ \rightarrow St \\ \rightarrow Bo \\ \rightarrow In' \\ \rightarrow Lo \\ \hline \end{array}$	→ VARCHAR an → BOOLEAN r → INTEGER → BIGINT e → DOUBLE	0	*
- LOCADALE - DATE	ccal Date → LocalDate → DATE	Mapping by Typ KNIME Type String Boolean value Number (integer Number (long) Number (double Local Date	r)			$\begin{array}{c} Map \\ \rightarrow St \\ \rightarrow Bo \\ \rightarrow In \\ \rightarrow Lo \\ \rightarrow Da \\ \rightarrow Lo \end{array}$	VARCHAR sn → BOOLEAN r → INTEGER → BIGINT = → DOUBLE hate → DATE	0	
.ocal Time → LocalTime → TIME	ccal Date → LocalDate → DATE ccal Time → LocalTime → TIME	Mapping by Typ KNIME Type String Boolean value Number (intege Number (long) Number (double .ocal Date .ocal Time	r)			Map → St → Bo → In → Lo → Lo → Lo → Lo	→ VARCHAR an → BOOLEAN er → INTEGER → BIGINT e → DOUBLE able → DATE ime → TIME		*
Jocar Jone → UocalJone → UA1C Local Time → LocalTime → TIME Local Date Time → LocalDateTime → TIMESTAMP	ccal Date → LocalDate → DATE ccal Time → LocalTime → TIME xcal Time → LocalDate Time → TIMESTAMP	Mapping by Typ KNIME Type String Boolean value Number (Integer Number (Iong) Number (double Local Date Local Time Local Date Time	r)			Mapp → St → Bo → In → Lo → Do → Lo → Lo → Lo	→ VARCHAR → VARCHAR an → BOOLEAN ar → INTEGER → BIGINT = → DOUBLE bate → DATE ime → TIME bateTime → TIMESTAMP		~

Reporting in KNIME

- Reporting in KNIME is done via a 3rd party application named BIRT (Business Intelligence Reporting Tool)
- Data is sent to BIRT from KNIME using special nodes.
- Reports in BIRT are constructed from report items, which may include images, tables, charts and labels.
- Reports may be generated in a variety of formats (html, pdf, pptx, xlsx, docx, ...)





Send Image / Data to Report

- Sends a data table or images to BIRT
- PNG and SVG are supported formats (see node description for details)
- Hint: The node label will be used to identify the data source in the reporting view
 Make sure to use fitting labels if you have more than one data source





Edit the Report

Open the workflow > Click the Report Editor button in the tool bar





Installing Extensions

- Install the KNIME Report Designer Extension to use BIRT
- Install extension by going to File -> Install KNIME Extension or via Drag & Drop from the KNIME Community Hub





Onen for Innovativ

KNIMF

Creating a Data App on KNIME Business Hub





Execute Workflow as Data App on KNIME Business Hub



© 2023 KNIME AG. All rights reserved.



Components to Produce Dashboard on Web Page





Exercise: 09_Deployment

- Write the clean customer data to an Excel file into the folder "data/temp"
- Write the full transaction data to the "Transactions.sqlite" database

Flow Variables



Goal of this Session

- What is a Flow Variable?
- Create a Flow Variable
- Use a Flow Variable as a parameter in the node settings
- Use a Configuration node to parameterize a Component
- Use a Widget node to enable interaction on a Data App page

Flow Variables: Usage Example

 Each month you need to produce a sales report for the most popular product




Flow Variables: Usage Example

- Each month I need to launch the Analytics Platform, aggregate the data to identify the most popular product, and update the Row Filter accordingly
- Or do I? Maybe Flow Variables can help...

Automatically Filter by Most Popular Product





Table Row to Variable

- Takes a table as input and converts the first row to Flow Variables
 - Column names -> Flow Variable names
 - Column values -> Flow Variable values
- Only the first row is transformed, additional rows are discarded



			Flo	ow Variables
In	Owne	Na	me	Value
0	4:6	s	Products	Private Investment
0	4:6	i	Count*(Age)	5296
0	4:6	s	RowID	Row4
0		s	knime.works	/Users/kathrinmelcher/knime-workspa.



Flow Variable Ports





Apply a Flow Variable (Button)

•••	Dialog - 0:274 - Row Filter		
Filter Criter Filter Criter Include rows by attribute value Exclude rows by attribute value Include rows by attribute value Include rows by number	Dialog - 0:274 - Row Filter ta Flow Variables Memory Policy Column value matching Column to test: S Products filter based on collection elements Matching criteria • use pattern matching Cold Investment Cold Investment	Variable Set	tings Product
Exclude rows by number Include rows by row ID Exclude rows by row ID	use range checking lower bound: upper bound: only missing values match		ОК
	OK Apply Cancel 🕅	The Flow Variable	button

÷

Cancel

Apply a Flow Variable (Advanced)

Filter Critera Flow Variables Memory Policy	_
▼ ? rowFilter	The Flow Variables tab
© ColumnName © Product ? include ? deepFiltering ? deepFiltering ? CaseSensitive © Pattern ? hasWildCards ? isRegExpr	List of available Flow Variables



Create a Flow Variable (Button)

 Filter Criter Include rows by attribute value Exclude rows by attribute value Include rows by number Exclude rows by number Include rows by row ID Exclude rows by row ID 	Dialog - 0:274 - Row Filter ia Flow Variables Memory Policy Column value matching Column to test: S Products + in filter based on collection elements Matching criteria • use pattern matching Cold Investment Cold Investm
	upper bound: only missing values match OK Apply Cancel



Create a Flow Variable (Advanced)

- Converting a setting value into a Flow Variable

Y 10	owFilter		
8	RowFilter_Typ	÷	
s	ColumnName	* New Variable	
?	include	\$	
?	deepFiltering	\$	Name of the ne
?	CaseSensitive	\$	Flow Variable
\$	Pattern	\$	
?	hasWildCards	\$	
?	isRegExpr	÷	
s ? ?	Pattern hasWildCards isRegExpr		



Sorter

- Sorts a table!
- Choice of ascending or descending
- Sort by multiple columns

Sorter

▶↓↑



Table "def	ault" – Rows: 5	Spec – Columns: 2
Row ID	S Products	Count*(Age)
Row4	Private Investment	5296
Row3	P+B Investment	5018
Row2	Gold Investment	3258
Row1	Fund Manager+	3143
Row0	CO Investment	1549



Variable Creator

- Allows to create flow variables of different types
- Click on "+ Add" to add a new variable and define a custom
 - Variable Name
 - Variable Value

	Va	ria	ble	Cre	ator
--	----	-----	-----	-----	------



Туре	Variable Name	Value	
i Integer 🗘	lower_bound	-10	↑ ↓ i
il Integer 🗘	upper_bound	10	
s String	default_column	Age	
			+ Add



Path Variables

- Special flow variable type to point to a file or folder
 - E.g. to control output location of a file
- A path type consists of three parts:
 - **Type**: Specifies the file system type e.g. local, relative, mountpoint, custome_url or connected.
 - Specifier: Optional string with additional file system specific information e.g. relative to which location (knime.workflow)
 - Path: Specifies the location within the file system

Туре	-Output locatio	on			Specifier
	Write to	Relative to	Current workflow	\bigcirc	
	File	/data/customer.csv		•	Browse
	Write options	Create missing folders	lf exists: 🧿 overwrite	🔵 append 🔘 fa	ail
201					Path

- Examples:
 - (LOCAL, , C:\Users\username\Desktop)
 - (RELATIVE, knime.workflow, file1.csv)
 - (MOUNTPOINT, MOUNTPOINT_NAME, /path/to/file1.csv)
 - (CONNECTED, amazon-s3:eu-west-1, /mybucket/file1.csv)



Create File/Folder Variables

- Creates one or multiple path flow variable(s) pointing to files / folders
- Inputs:
 - Base location
 - Flow variable name(s)
 - Value (file name or path relative to base location)
 - File extension (optional)
- Output variables can be used to control the output location in writer nodes.

Create for	Relative to V	urrent workflow 🔍		
Folder	//data/temp/products		~	Browse
File/Folder	variables			
Variable filePath	Base location	Value	File exte	+ Add variable
				Remove variable
🛕 The "pa	th_values" parameter	is controlled by a variable	e.	
	e/Folder	ОК	Apply Cano	el 🕜
ate Fil				



Example: Add Execution Date to File Name





Configuration Nodes for Variable Creation and Output







Configuration Node Configuration

Use Configuration nodes to create Flow Variables



Control Flow Variables	Job Manager Selection	Memory Policy	
Label:	Select product:		
Description:	Select the product, whi ded.	ch should be inclu	
Parameter/Variable Name:	product_select		
Selection Type:	Dropdown		
Lock Column:			
Default Column:	S Products	0	
Default Value:	Gold Investment		
Limit number of visible options:			
Number of visible options:		10 (



Simple Configuration of Component



- Double click a component to configure it
- For use on the KNIME Business Hub as a Data App, replace Configuration nodes with Widget nodes

<u>.</u>	Flow Variables	Memory Policy	Job Manager	Selection	
Coloct p	aduct.				
select p	ouuct:				
Privat	te Investment 🕔				
Privat	e Investment				
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	nvestment				
P+BI					
P +B I Gold I	nvestment				
P +B I Gold I CO In	nvestment vestment				



Multiple Selection Configuration Node

_	Control	variables Memory Foncy		Control Prow Var	indies Memory Foncy
L	abel:	Select the products to include	► ? de	efaultValue	
	Description:	Enter Description	b [®] hi	ideInDialog	
			s" pa	arameterName	\$
			s' la	ibel	
F	arameter/Variable Name:	multiple-selection	s" de	escription	\$\$\$
5	election Type:	Twinlist	b [®] re	equired	\$
			s" flo	owVariableName	\$
F	ossible Choices:	Product 1 Product 2	sů po	ossible_choices	st AggregatedValues
		Product 3	s° ty	/pe	
			b [•] lir	mit_number_visible_options	
	Default Values:	Product 1	[i " nı	umber_visible_options	\$
		Product 2			
		Product 5			
	imit number of visible options:				
ſ	lumber of visible options:	10	List of P	ossible Choic	es can
le Selection	1		be contro	olled by an ar	cay flow
figuration	ОК	Apply Cancel			
				variable	

1.1



Array Variables

- An array flow variable is a flow variable with many values
- Can be used to control setting options with more than one value, e.g.
 - Include / exclude of columns
 - Define list of possible values
 - Control the variable names and values in the Create File / Folder Variables node
- How to create an array flow variable?
 - Step 1: Create a collection cell (e.g. Create Collection Column node or GroupBy node)
 - Step 2: Use the Table Row to Variable node to create the array flow variable based on the collection cell





Example: Creating an Array Flow Variable





Components

- Encapsulate a functionality for reuse and sharing
- Main features:
 - Local Flow Variable scope
 - Configurable via Configuration nodes
- Key to advanced functionality in KNIME products
 - Component corresponds to a KNIME Data App page
 - Configurations on a Data App page are defined using Widge
 - Can be shared via KNIME Community Hub





Component Description

▲ *6: 07. Flow Variables - solution 🔺 *6:11:0 - Filter by Product (select product category) 🛛 😐 🗖	€ Description Ω		
Value Selection Component Input	Description		
< v	Ponts Sink In Port # 5Source Tription: Data containing records for all products Out Port #1 Name Filtered data Description: Data containing records for the selected product	▲ 16 07 Flow Variables - solution 13	Description 10 Filter by Product This component filters the data by the selected product name Dialog Options Select product: The product name to use for filtering

- Double click a component to configure it
- For use on the KNIME Business Hub as Data App, replace Configuration nodes with Widget nodes



Configure Component Ports

ilter by pro	duct				
	Configure	F6	i i		
0	Execute	F7			
8	Execute and Open Views	Shift+F10			
0	Cancel	F9			
圜	Reset	F8			
=	Edit Node Description	Alt+F2			
6	New Workflow Annotation				
-	Connect selected nodes	Ctrl+L			
20	Disconnect selected nodes	Ctrl+Shift+L			
	Create Metanode				
	Create Component		L .		
22	Component	>		Open	
Q	Interactive View: Filter by product		5	Expand	
₽?	Compare Nodes			Setup	
	Show Flow Variable Ports			Convert to Metanode	
ot	Cut		2	Share	
	Copy			Update Link	Ctrl+Alt+U
1	Paste			Disconnect Link	
2	Undo			Change Link Type	
3	Redo			Select in Explorer	
×	Delete				
E	Port 1				

- Add input and output ports to metanodes/components
- Remove ports to adapt to changes after creation of the metanode/component

Specifiy the name of the n Component Name: Filte	ode and define the numl r by product	ber and type of the desired in	and out ports.	Port Type:	Data Data Flowvariable PMML
In Ports: in_1 (Data)	Add Remove Up Down	Out Ports: out_1 (Data)	Add Remove Up Down		Database Query AWS Comprehend Connection AWSConnection DB Data DB Session Distance Measure FilterDefinition Gradient Boosting Model H2O Context
		• Fin	sh Canc	e	H2O Frame H2O Model Image KnimeConnection M0JO Outlier Python Python Regression Tree



Passing Variables from Components

- Flow Variables are -by default only available locally inside the component
- Configure the component input/output to pass Flow Variables from/to outside the component



orniguration	Descriptions	Flow Variables	Job Manager Selection	Memory Policy		
		Choo	se variables from workf	ow to be visible outside t	the Component	
			Manual Selection	O Wildcard/Regex Se	lection	
Exclude				[Include -		 ٦.
T Filt	'er			T Filte	r	
s° sele	cted product				No variables in this list	
s° sele	cted_product (column)		>		
				>>		
				<		
				~		
⊖ Enfo	rce exclusion			Enforce	e inclusion	
						-



What is a Shared Component?

- Components can be saved in your KNIME workspace for later reuse
- To do this, simply right-click any component and select "Share..."
- Shared components are read-only instances of a component
- Public Shared Components are available on the EXAMPLES Server and on the KNIME Community Hub



How can you Edit a Shared Component?

- Components can be edited using the Component Editor, similar to workflows
- To edit a component using the Component Editor, double-click the component in its location in the KNIME Explorer
- To ensure components are executable when opened in the Component Editor, choose the option to "Include input data with component" when sharing it





How can you Use a Shared Component?

- To use a Shared Component, drag and drop it to your workflow editor
- Instances of Shared Components can be updated either manually or when the workflow is opened
- A Shared Component can also be unlinked from its original location, which makes it editable in the workflow directly
- Update Shared Components by overwriting them







Configuration Dialog Layout



 Click layout button when inside component to modify the order of the setting options in configuration window of the component

t for the KNIME WebPortal and the composite view. Ier of the contained configuration nodes for the configuration	n dialog of the component.		-
Composite View Layout Advanced Composite View Layout	Configuration Dialog Layout		_
C	Column Selection Configuration Node 24 Configuration columns		
	Integer Input Node 28 Configuration min		
			i.
	for the KNIME WebPortal and the composite view. ler of the contained configuration nodes for the configuration Composite View Layout Advanced Composite View Layout c	tor the KNIME WebPortal and the composite view. ler of the contained configuration nodes for the configuration dialog of the component. Composite View Layout Advanced Composite View Layout Configuration Dialog Layout Column Selection Configuration Node 24 Configuration columns Integer Input Node 28 Configuration min	ter of the KNIME WebPortal and the composite view. ter of the contained configuration nodes for the configuration dialog of the component. Composite View Layout Advanced Composite View Layout Configuration Dialog Layout Column Selection Configuration Node 24 Configuration columns Integer Input Node 28 Configuration min

lumeric columns:			
age 🗸 🗸			
lin	10 ‡		Change
1ax			Change
	40 🌲		



Breakpoint

- Stops execution of a workflow branch
- Useful to stop the execution of a component and provide a custom error message
- Execution stops based on the selected condition:
 - Empty table
 - Active/Inactive branch
 - Flow Variable value



🛕 Dialog File	- 2:117:0:143:1	38 - Breakpoi	int (Fail if	Number Col.	. –		×
Options	Flow Variables	Job Manager	Selection	Memory Policy	1		
		🖂 Br	eakpoint E	nabled			
Breal	kpoint active for: mpty table) ;	active branch Select Variable	inactiv	e branch) 💿 er Columns 🗸	variable ma	atches value	
		Enter Varia	able Value:	0			
Custom message No numeric column available							
	OK		Apply	Can	el	0	



Widget Nodes for Variable Creation and Output







Simple Configuration of Component



	negation name in a contract of the track of
xcludes	Includes
Private Investment CO Investment Fund Manager+	> P+B Investment Gold Investment



Summary: Flow Variables

- Flow Variables are workflow parameters used to overwrite existing node settings
- A Flow Variable is carried along workflow branches (parallel branches don't share local Flow Variables)
- Flow Variables can be of type String, Integer, Double, Boolean, Long, Paths and Array
- Flow Variables can be created
 - 1. in the "Flow Variables" tab of any node
 - 2. using specific nodes, e.g. Table Row to Variable, Variable Creator
 - 3. using Configuration and Widget nodes



KNIMF

Exercise: 10_Flow_Variables

- Activity I: Filter the customer data to
 - Customers of the "Gold Investment" product
 - Customers of the most common product in the data
- Activity II: Create a component that allows a user to choose an investment product and filter the data by that product
- Activity III (optional): Create a path variable that automatically has the current execution date in the file name and write the filtered table into a CSV file
- Activity IV (optional): Create a component that allows to select multiple products out of all available products, using a flow variable of type array

KNIME Learning Paths

From level L1 to level L4 for various professional profiles



Self-paced courses: videos and exercises at your own pace and for free

Instructor led courses: Scheduled sessions and guided exercises in paid courses



knime.com/knime-courses



Workflow Control Loops, Switches, Try-Catch

Workflow Control Structures

	V Workflow Control
	Automation
	V Variables
	Breakpoint
	T Counting Loop Start
	Chunk Loop Start
	T Column List Loon Start
	Generic Loop Start
J V V	Table Row To Variable Loop Start
	loop End
	Veriable Condition Loop End
	Variable Loop End
	I Group Loop Start
	Interval Loop Start
WC	Loop End (Column Append)
	Recursive Loop End
	C Recursive Loop Start
	v to Switches
	IF Switch
	>IF End IF
	CASE Switch Start
	CASE Switch End
	Empty Table Switch
	ir Java IF (Table)
	✓ S Error Handling
	Catch Errors (DB Ports)
	Catch Errors (Data Ports)
	Catch Errors (Generic Ports)
	Catch Errors (Var Ports)
	Try (Data Ports)
	Try (Variable Ports)
	Active Branch Inverter
	> 📴 Meta Nodes

- Loops
 - Iterate over a workflow snippet with variable inputs.
- Switches
 - Direct the path of a workflow by selectively executing one or more workflow branches.
- Try-Catch
 - Handle workflow branches that may fail in execution when you don't know about this before executing



The Loop Block

- A loop block is defined by the appropriate loop start and loop end nodes.
- Loop body = the nodes in between (including the side branches).



KNIME
Group Loop Start

- Similar to GroupBy except without aggregation tab.
- Each iteration of the loop passes the next group of rows.
- You can implement an aggregation task. It can be anything from a complex calculation to updating a database.

Exclude	
S MaritalStatus S Gender EstimatedYearlyIncome NumberOfContracts Age Target Available401K CustomerValueSegment ChurnScore CallActivity	S Products







Example: Writing Aggregated Files

- Group data by specific column values
- Iterate over all groups of data
- Create an appropriate file name
- Write grouped data to tables with new file name





Create File/Folder Variables

- Creates one or multiple path flow variable(s) pointing to files / folders
- Inputs:
 - Base location
 - Flow variable name(s)
 - Value (file name or path relative to base location)
 - File extension (optional)
- Output variables can be used to control the output location in writer nodes.

Create for	Relative to V	urrent workflow 🔍		
Folder	//data/temp/products		~	Browse
File/Folder	variables			
Variable filePath	Base location	Value	File exte	+ Add variable
				Remove variable
🛕 The "pa	th_values" parameter	is controlled by a variable	e.	
	e/Folder	ОК	Apply Cano	el 🕜
ate Fil				



Example: Writing Multiple Excel Sheets





List Files/Folders

- List all files in a directory
- Restrict to:
 - Top level directory (i.e. not recursive)
 - Specific file extensions
 - Matching name patterns (regex or wildcard)
- Provide file references as a table of URLs and absolute paths

List Files/Folders



ettings C	au Vasiables - 3-6 au	Colorkan Managara			
Read from Mode Folder	Relative to Files in folder () Files in folder () Selected 5 of 5 parent folder der identifier column	Current workflow	Filter options In Specify extensio	clude subfolders file on	
¢		 ▲ Filter options File filter options ■ File extension(s) csv □ Case sensitive □ File name * □ Case sensitive □ Include hidden files 	 ⊙ Wildcard ○ Regular e 	expression	>
		Folder filter options	Wildcard Regular e	expression	



Example: Reading Many Excel Sheets

- List all sheet names of an Excel file
- Convert sheet name into a flow variable (1 sheet name per iteration)
- In each iteration, read the spreadsheet with the current sheet name
- Close the loop and collect the results





Table Row to Variable Loop Start

- Similar to the Table Row to Variable node
- Each iteration of the loop converts the next row of the input table into Flow Variables
- Injects variables into other nodes to re-execute subflows with a progression of settings



Settings Flow Variables Job Manager Selection	
Missing values Handling	
⊖ Fail	
Use defaults if available	
⊖ Omit	
Defaults	
String missing	
Boolean false 🗸	
Integer 0 🜩	
Double 0.0 🖨	
<i>Filter</i> No columns in this list	> P Path
	>>> < <
Inforce exclusion	>> <



Loop End

- Can be used to end of a loop
- Collects the results of the different iterations by row-wise concatenation of the incoming tables
- Provides options to:
 - Add a column with the iteration number
 - Propagate modified loop variables
 - Allow variable column types
 - Allow changing table specifications

Loop End



Standard settings	Flow Variables Job M	Manager Selection	Memory Policy		
	Row key po Generat Unique r Leave ru Add itera Propagat Jignore er Allow var Allow cha	licy te new row IDs row IDs by appendir ow IDs unmodified tion column we modified loop vari mpty input tables iable column types inging table specifica	ng a suffix ables ations		
	OK	Apply	Cancel	2	



Switches

- A switch allows you to selectively activate branches of a workflow
- Inactive branches are marked with a red x on their output ports.
 Inactive nodes propagate down stream.





Single Selection Configuration

- Configuration: Select single value from list of Strings
- Returns selection as string type Flow Variable
- Choose between different layout options (dropdown, radio buttons...)



Control	Flow Variables	Job Ma	anager Selection	Memory Policy	
Label:			Select plot type	:	
Descript	ion:				
_					
Paramet	er/Variable Nam	e:	plot_type		
Selection	n Type:		Dropdown		~
Possible	Choices:		bar		
			scatter		
Default	Value:		bar scatter		
			Jearcer		
Limit nur	mber of visible op	otions:			
Number	of visible options	s:			10
			L		

Rule Engine/Rule Engine Variable

- Defines custom logic for using simple rules.
- Rules like: <Antecedent> =>
 <Consequence>
 (1=1 => "true")
 - (1=1 => "true")
- May be used in Flow Variables or tables
- Easiest way to encode logic for switches





If Switch

- Controls which branches of your workflow are active programmatically
- Controlled with a Flow Variable, setting the value to the literal Strings: "top", "bottom", "both"
- May be used in Flow Variables or tables (different nodes)









Case Switch Start & End

- Similar to If-Switch: Takes data from single input port and passes it to the active output port
- Nodes connected to inactive branches are not executed





Case Switch Start & End

- Case Switch Start
 - Add an input port with a specific type (e.g., Data)
 - Two output ports are also added
 - Additional output ports can be added
- Case Switch End
 - Add an output port with a specific type (e.g., Data)
 - Two input ports are also added
 - Additional input ports can be added

- Configure via node dialog, or pass port index as Flow Variable
 - From the top, 0, 1, 2, ... (however many ports there are)



The Difference between Loops and Switches

Loops

- The Loop Start is connected to the Loop End node; they form a pair.
- A loop iterates over a workflow part.

Switches

 A Switch Start can be used without a corresponding Switch End. They can also be combined.



Try-Catch

- A way to catch errors in workflows
- Useful when it is hard to know if a node will execute (for example, when reading from a Google Sheet)
- KNIME tries to execute the nodes, but if it fails will fall back to an alternative branch

Regular Execution





Alternative Execution



Streaming

- Standard execution: Node by node. The node processes all data, finishes, then passes the data to the next node, etc.
- Streaming: Nodes executed concurrently, each nodes passes data to the next as soon as it is available, i.e. before node is fully executed
 - Faster execution, esp. for reading/preprocessing data
- Install KNIME Streaming Execution (Beta) extension
- Create Component -> Configure -> Job Manager Selection -> Simple Streaming
 - Not available for all nodes (show in node repository)
 - Can only execute entire metanode, not individual nodes
 - Intermediate results not available since nothing is cached

89



Streaming





Exercise: 11_Workflow_Control, Activity I

Goal: Build a loop that will create an Excel file with separate Excel sheets for the records of different products.

- Read the table CurrentDetailData.table (Table Reader node)
- Start a loop that handles the records for the different products in separate iterations (Group Loop Start node)
- For each product write one Excel sheet into a single Excel file (Excel Writer node)
- Close and execute the loop (Variable Loop End node)



Exercise: 11_Workflow_Control, Activity II

Goal: Create a loop that reads and concatenates all the sheets in an Excel file.

- Create a table that contains all sheet names of the Excel file created in Activity I (Read Excel Sheet Names node)
- Start a loop that iterates over the sheet names (Table Row to Variable Loop Start node)
- Read the Excel sheet with the sheet name in the current iteration (Excel Reader node)
- Close the loop and concatenate the tables from the different iterations (Loop End node)



Exercise: 11_Workflow_Control, Activity III

- Extend the workflow below with a switch that only creates one type of visualization
 - Create a Single Selection Configuration node with the possible values "scatter" and "bar"
 - Use the CASE Switch Data (Start) that activates the top or the middle branch depending on the selection scatter/bar (Use the "...(index)" flow variable to define the active port)
 - Combine the outputs of the two branches with the CASE Switch Data (End) node





Instructor-Led Courses: Schedule 2023

Sep 11 - 16	L1	L1-DW KNIME Analytics Platform for Data Wranglers: Basics	
Sep 18 - 22	L2	L2-DW KNIME Analytics Platform for Data Wranglers: Advanced	Data
Oct 23 - 27	L3	L3-DA Productionizing Data Apps*	Analysts
Nov 6 - 10	L4	L4-DV Low Code Data Extraction and Visualization*	

Jun 19 - 23	L1	L1-DS KNIME Analytics Platform for Data Scientists: Basics	
Jun 26 - 30	L2	L2-DS KNIME Analytics Platform for Data Scientists: Advanced	
Jul 10 - 14	L3	L3-CD Continuous Deployment and MLOps*	
Jul 17 - 21	L4	L4-ML Introduction to Machine Learning Algorithms	Data Scientists
Jul 24 - 28	L4	L4-DL Introduction to Neural Networks and Deep Learning	Sciencists
Aug 21 - 25	L4	L4-TS Introduction to Time Series Analysis*	
Aug 28 - Sep 1	L4	L4-TP Introduction to Text Processing	

Sep 11 - 16	L1	L1-DW KNIME Analytics Platform for Data Wranglers: Basics	
Sep 18 - 22	L2	L2-DW KNIME Analytics Platform for Data Wranglers: Advanced	Data
Oct 9 - 13	L3	L3-DE Productionizing Data Engineering Applications*	Engineers
Oct 16 - 20	L4	L4-DE Best Practices for Data Engineering*	





Styling EXCEL Tables



Styling EXCEL tables

Standard written table

•••	 AutoSa 		alter v v v ≠ ≞itest	_conti — Saved to my Mac ~			् प ्
Home	Insert	Draw Page I	ayout Formulas Data.	Review View		B Share	し Comments
к4	¥ ^	√ Jx					
	Α	В	С	D	E		F
1	Year	Quarter	Store - no CC	Store - with CC	OnlineStore		
2	2019	1	36862,74	66775,81	114196,84	1	
3	2019	2	38059,65	70483,79	113399,81	Ĺ	
4	2019	3	48149,06	76791,58	96116,79)	
5	2019	4	47220,13	61563,41	105625,31	Ľ	
6							
7							
4 1-	default	+					

Styled table

••	AutoSi	ave 🔵 orr) 🏠 [ਜ਼ ਦਿ° ਨ × ੴ ਵ 🖻 test	_conti_2 — Saved to my Mac		م © •
Home	Insert	Draw Page	Layout Formulas Data	Review View	ß	Share U Comments
	A	В	С	D	E	F
1	Year	Quarter	Store - no CC	Store - with CC	OnlineStore	
2	2019	1	36862,74	66775,81	114196,84	
3	2019	2	38059,65	70483,79	113399,81	
4	2019	3	48149,06	76791,58	96116,79	
5	2019	4	47220,13	61563,41	105625,31	
6						
7						
4 >	default	+				

Continental Nodes for KNIME XLS Formatter Nodes



Additional tag table - the key to a styled table:





XLS Control Table Creator

- Kick off node to your styled table
- Transforms input table into an XLS Control Table
- Column names => A, B, C, ….
- Row IDs => 1, 2, 3, …
- The unpivoted table is a great base to transform values into tags, e.g. with the Rule Engine node

	Options	Flow Variables	Job Manager Selectio	n Memory Policy	
Operatio	n Type (auto	matically set base	ed on the provided inpu	t table)	
	• from arb	itrary input table t	to XLS Control Table (wi	de or long/unpivoted)	
	from lon	g/unpivoted layou	ut to wide XLS Control ٦	able	
Shift Rov	vs Option				
		write c	olumn header to first n	w	
	unpiv	e Options ot result table (fo	r easier post-processir	ig and re-pivoting)	
	unpiv	e Options rot result table (fo 🖌 add ac	r easier post-processir dditional header columi	ig and re-pivoting) 15	
Contradi	able Structur unpiv	e Options rot result table (fo	r easier post-processir dditional header column peration Type 'long to t	ig and re-pivoting) 15 wide'	
Contradi how to	ction Resolu	e Options rot result table (fo add ac tion Strategy at O ontradicting inforr	r easier post-processir dditional header column peration Type 'long to m nation? [fail]	ig and re-pivoting) 15 vide'	0
Contradi how to	ction Resolu	e Options vot result table (fo vot result table add vot add add add add vot add add add add vot add add add vot add add add vot add add add vot add add add add vot add add add add vot add add add add vot add add add add add add add add add ad	r easier post-processir dditional header column peration Type 'long to mation? fail	ig and re-pivoting) ¹⁵ vide'	0

XLS Control Table Generator





XLS Background Colorizer

- Changes the background of a cell into a static color and/or pattern fill
- Select between
 - Assigning a color to a tag value
 - Using a table with RGB values

XLS Background Colorizer



) 🔵 Dialog - 2:319 - XLS	Background Colorizer (header: ye
Options	Flow Variables
Control Table Style	
o standard tags	direct color codes in RGB format
Tag Selection applies to tag (single	tag only) header
Background Color	
Change color?	color 🦲 Change
Pattern Fill	
pattern fill	unmodified ᅌ
Change color?	color Change
ОК Ар	ply Cancel



XLS Conditional Formatter

- Changes the cells' background according to their numerical value
- Allows you to optionally define a mid scale point



👂 🔵 Dialo	g - 2:329 - XL	S Conditional Formatter						
Optic	ns	Flow Variables						
Tag Selection								
applies to tag (single tag only) value								
Conditional	Formatting Se	ettings						
	🗹 Mid sca	le point needed?						
min 0.0 Å		min color Change						
	0.0							
mid	0.5 🗘	mid color 🦰 Change						
max	10 0	max color Change						
	1.0							
ОК	Apply	Cancel						



XLS Border Formatter

- Changes the borders of a given range specified by tags
- Option to use boarders inside the specified range
 - Inner vertical
 - Inner horizontal



applies to tag (single tag only) head	der applies to all tags
Border Style and Color	
border style	normal
Change border color?	border color Change
Outer Border Settings	
	C top
🗌 left	ight
	🗸 bottom
nner Border Settings	
inner vertical	inner horizontal



XLS Formatter (apply)

- Applies the chained commands from the XLS Formatter nodes
- Important:
 - The table must be written before with Excel Writer node
 - The input file may not contain any formatting yet

Excel Writer	XLS Formatter
Excerviter	(apply)
► = x	/ 🖸
Write standard table	Write styled table
	/
	/
/	(

ttings Flow	Variables			
Source Read from	Current workflow			
ile [.	.//data/Revenue_2019.xlsx	~	Browse	 •
Destination — Write to	Relative to V Current workflow V			
ile	//data/Revenue_2019_styled.xlsx	~	Browse	 V.
Write options	□ Create missing folders If exists:			
] open outpu	It file after execution			

102

KNIME

Other Important Nodes of the Extension



By default the styling is always applied to the first sheet. Change this with the XLS Sheet Selector node.

Want to style more than one sheet? Merge different sheet stylings with the XLS Format Merger node.

XLS Sheet Selector



XLS Format Merger







Want to add a header to your table? Merge multiple cells with the XLS Cell Merger node.



Tips & Tricks for the Continental Extension

- How can you automate creating a tag table?
 - Handle changing number of rows
 - Handle changing number of columns
- How can you add empty rows between different tables?

Hon	ne Insert Draw	Page Lavor	t Formulas	Data Revie	w View		A Share	Comment
1101		ruge Luyet	it ronnaid.	butu nevi	in them		E ondre	
L28	$ \times \sqrt{f_x}$							
	А	В	С	D	E	F	G	Н
1		Bas	ic Repor	t Example	Mai 2019			
2								
3	Calendar Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
4	CW 18			2158.99	2486.56	2756.36	5486.35	
5	CW 19	2196.47	2963.47	2348.57	2786.57	2873.56	3947.56	
6	CW 20	2964.87	2145.78	3120.86	3452.99	3245.58	4279.26	
7	CW 21	2115.78	2657.23	2765.36	2275.96	1458.65	4823.45	
8	CW 22	2678.36	1268.46	2192.78	2571.45	3156.47		
9								
10								
11								
12								
4	default +							



Tips & Tricks for the Continental extension







Tips & Tricks for the Continental extension

		• •			Control Table - 5:	2260 - XLS Control Tal	ole Generator (Create	e unpivoted)			
		File Hilite	Navigation	View							
le Creato	Add				Table "default" - Row	s: 56 Spec – Colum	ns: 8 Properties	Flow Varia	ables		
he oreator	Auu	Row ID	S Cell	S Column	S Column (comparable)	Column (number)	S Column name	Row	S RowID	S Value	
		Row0	A1	A	00A	1	column1	1	Row0	Basic Report Example Mai 2019	
⊞. ►		Row1	B1	В	00B	2	column2	1	Row0	?	
		Row2	C1	с	00C	3	column3	1	Row0	?	
		Row3	D1	D	00D	4	column4	1	Row0	?	ormatter
		Row4	E1	E	00E	5	column5	1	Row0	?	(vla
		Row5	F1	F	00F	6	column6	1	Row0	?	
Header		Row6	G1	G	00G	7	column7	1	Row0	?	
		Row7	A2	A	00A	1	column1	2	Empty 0	?	
		Row8	B2	В	00B	2	column2	2	Empty 0	?	
	Extract	Row9	C2	с	00C	3	column3	2	Empty 0	?	
	ne	Row10	D2	D	00D	4	column4	2	Empty 0		ach cell one row
	pe	Row11	E2	E	00E	5	column5	2	Empty 0	?	
		Row12	F2	F	00F	6	column6	2	Empty 0	[?] with r	ow and column
		Row13	G2	G	00G	7	column7	2	Empty 0	?	
		Row14	A3	A	00A	1	column1	3	Row0_dup	Calendar Wee num	per and Cell ID
	-	Row15	B3	В	00B	2	column2	3	Row0_dup	Monday	
		Row16	C3	с	00C	3	column3	3	Row0_dup	Tuesday	
		Row17	D3	D	00D	4	column4	3	Row0_dup	Wednesday	
		Row18	E3	E	00E	5	column5	3	Row0_dup	Thursday	
		Row19	F3	F	00F	6	column6	3	Row0_dup	Friday	
		Row20	G3	G	00G	7	column7	3	Row0_dup	Saturday	
		Row21	A4	A	00A	1	column1	4	Row1	CW 18	
		Row22	B4	В	00B	2	column2	4	Row1	?	
		Row23	C4	С	00C	3	column3	4	Row1	?	
		Row24	D4	D	00D	4	column4	4	Row1	2158.99	
		Row25	E4	E	00E	5	column5	4	Row1	2486.56	
		Row26	F4	F	00F	6	column6	4	Row1	2756.36	
		Row27	G4	G	00G	7	column7	4	Row1	5486.35	
					~~.		· ·	-		QU.1.0	



Tips & Tricks for the Continental extension

•	Dialog - 5:2261 - Rule Engine (Replace values) Rule Editor Flow Variables Job Manager Selection Memory Policy	Define rules based on the row and the column numbers
Column List ROWID ROWINDEX ROWCOUNT S Cell S Column (comparable, I Column (number) S Column name I Row S RowUD S Value	Category Description All ○ Function • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? < ? • ? > ? • ? AND ? • ? NA? • ? AND ? • ? NA? • ? AND ? • ? NO? • State • MISSING ? • NO ? • Pacesoin • ? 2 // \$double column name\$ > 5.0 => "large" ? 3 // \$string column name\$ LIKE "wblue*" => "small and blue" ? 4 // TRUE => "default outcome" § \$Strow\$= 1 => "header" • § \$Strow\$= 2 => "empty" § \$Strow\$= 4 => "border" § \$Strow\$=4 => "border" § \$Strow\$=4 => "border" </th <th>Excel Writer Image: State of the state of th</th>	Excel Writer Image: State of the state of th
0	Append Column: prediction S Replace Column: S Value	numbers



Excel Cell Updater

- Updates values of cells at specified address in an Excel file
 - While retaining the cell format



Provide an input table of cell addresses and values to be replaced

Example

Address	String	Integer
A5	Ok	?
5:96	?	50
OZ23914	?	?

- Missing values are denoted by "?"
- Cell A5 is replaced with a string "OK"
- Cell 5-th row, 96-th column is replaced with an integer 50
- Cell OZ23914 is cleared


Excel Cell Updater



Table with replacement values

Row ID	S Address	S String	S Integer
Row0	H13	yes	?
Row1	H16	yes	?

1	Α	В	С	D	E	F	G	н	
1	age	job	marital	education	default	balance	housing	loan	con
2	58	managem	married	tertiary	no	2143	yes	no	unk
3	44	technician	single	secondary	no	29	yes	no	unk
4	33	entrepren	married	secondary	no	2	yes	yes	unk
5	47	blue-colla	married	unknown	no	1506	yes	no	unk
6	33	unknown	single	unknown	no	1	no	no	unk
7	35	managem	married	tertiary	no	231	yes	no	unk
8	28	managem	single	tertiary	no	447	yes	yes	unk
9	42	entrepren	divorced	tertiary	yes	2	yes	no	unk
10	58	retired	married	primary	no	121	yes	no	unk
11	43	techniciar	single	secondary	no	593	yes	no	unk
12	41	admin.	divorced	secondary	no	270	yes	no	unk
13	29	admin.	single	secondary	no	390	no	no	unk
14	53	technician	married	secondary	no	6	yes	no	unk
15	58	technician	married	unknown	no	71	yes	no	unk
16	57	services	married	secondary	no	162	no	no	unk

Updated Origin Excentrate file



is preserved



Tips & Tricks for the Continental extension





Installing Extensions

- Install the Continental Nodes for KNIME extension
- Install extension by going to File -> Install KNIME Extension or via Drag & Drop from the KNIME Community Hub

KNIME	File Edit View Node	Search		0	hub.knime.com	C	O 🛆 🗗 +	
	New	% N %S		Q Search work	flows, nodes and more	?	: Sign in	
al Nal:	Save As		WHINE Hub . Continental WHINE	Extension Developer - Extensions	Continental Nodeo for VNIME			
IME Explore	Save All	<u> </u>	Knime nuo 7 continentai_knime	Contension Contension () Extensions	Concernal Hours for Knime			
3 🖻 🚸	Close All	ት መ	- Extension					
My-KNIME	Recent Workflows	► 5	Continental	Nodes for KNI	ME 🖉		G	
LOCAL (Lo	👜 Print		v 1.2.0			m	♡ 0	
	Import KNIME Workflow.	Available Software						
	Export KNIME Workflow.	Check the items that you wish to install.	The Continental Nodes for reporting canabilities inter	KNIME Extension provides da ided for business users of the	ata processing and KNIME Analytics	Legal		
	Switch Workspace	type filter text Select All	Platform. This extension c	omprises the XLS Formatting	nodes.	Copyright (c) 25 0 C Corporation. All r © License		
	Preferences	Name Version Deselect All P UID KNUME & Extensions P UID KNUME Flip Data Extensions	⊕ Included nodes ∞{	s Related workflows	● ● ● □ • ■ 0 100% ■ 8 ∞ A P O	KNIME Analytics Platforr		e Q
	Fimport Preferences	Image: Second Seco			A KNIME Explorer 12	- D 🛆 3: KNIME, project (8)		- C A Description SS - C
	Install KNIME Extensions	HUL KNIME Community Extensions - Cheminformatics HUL KNIME Community Extensions - Image Processing and Analysis HUL KNIME Community Extensions - Other	Network Component Splitter	This node analyzes a between nodes for un				Row Filter
	Update KNIME	b00 KNIME Community Extensions Sources		components. It expec Community Nodes > Contin	COCAL (Local Workspace) Example Workfloera	My first Workflow		The node allows for row filtering
	Restart				Customer Intelligence Retail	File Reader	Row Filter Column Filter Table Writer	include or exclude: certain ranges (by
	Exit				► TSocial Media	D	→ <mark>⇒</mark> ≻──→ <mark>₽</mark>	and rows with a certain value in a selectable column (attribute) Below are
		Details			Recommended Nodes Community	read adult.cav	keep only records remove gender Write table	the steps on how to configure the node in its configuration dialog. Note: The node
					GroupBy 14%			doesn't change the domain of the data table. I. e. the upper and lower bounds or
		Show only the latest versions of available software.			Concatenate 4%			A KNIME Hub Search 32
		Group items by category What is already installed?			Reference Row Filter 3% F[5] String Manipulation 3%			Search workflows, nodes, and more
		Show only software applicable to target environment			figit Math Formula 2%	a a St Outine 13	Console A Node Monitor 23	****
					A Node Repository		Node: Row Filter (3:2)	
					▶ 10 ▶ Manipulation	3333	State: EXECUTED	
					► Q Views ► Interviews ► E De		Port Output Port O D Load data	ration or marity status occupation relations?
					Other Data Types Other Data Types O Structured Data		Row1 50 Self-emp-not-inc 83311 Bachelors 13 Row2 38 Private 215646 HS-grad 9	Married-clv-spouse Exec-managerial Husband Divorced Handlers-cleaners Not-in-fa
		< Back Next > Cancel Finish			Gripting Gots & Services Community Nodes		Row3 53 Private 234721 11th 7 Row5 37 Private 284582 Masters 14 Row7 52 SetTemp-not-inc 20642 HS-grad 9 Row7 52 SetTemp-not-inc 20642 HS-grad 9	Married-civ-spouse Handlers-cleaners Husband Married-civ-spouse Exec-managerial Wife Married-civ-spouse Exec-managerial Husband
								1



Exercise: 12 Styling Excel Tables

- Read the files: TagTable.table and RevenuePerQuarter.table and have a look at the tables.
- Use the XLS Formatter node to change the styling of the RevenuePerQuarter.table table to look like this, below:

Hint: Use the following nodes:

- XLS Control Table Generator
- XLS Background Colorizer
- XLS Border Formatter
- XLS Row and Columns Sizer (optional)

	AutoSave) n n n	რ ∽	conti_2 ~		Q © v
Home	Insert Dr	aw Page Layout	Formulas Data Review	View	🖻 Share 🖓 C	omments
01	‡ × ∨	fx				Ŧ
	А	В	С	D	E	F
1	Year	Quarter	Store - no CC	Store - with CC	OnlineStore	
2	2019	1	36862,74	66775,81	114196,84	
3	2019	2	38059,65	70483,79	113399,81	
4	2019	3	48149,06	76791,58	96116,79	
5	2019	4	47220,13	61563,41	105625,31	
6						
7						
• •	default	+				
					· +	267%



Introduction to Data Science



Churn Prediction



CRM System Data about your customer

- Demographics
- Behavior
- Revenues



Model



- Churn Prediction
- Upselling Likelihood
- Product Propensity /NBO
- Campaign Management
- Customer Segmentation

• • • •



Customer Segmentation



CRM System Data about your customer

- Demographics
- Behavior
- Revenues



Model



- Churn Prediction
- Upselling Likelihood
- Product Propensity /NBO
- Campaign Management
- Customer Segmentation

• • • • •



Demand Prediction



How many taxis do I need in NYC on Wednesday at 12:00?





Recommendation Engines / Market Basket Analysis



Sentiment Analysis



Samsung

Samsung Galaxy S7 Edge G935A 32GB Unlocked - Gold Platinum

☆☆☆☆☆ Beautiful phone from a wonderful seller!

By . on May 29, 2017 Color: Gold | Verified Purchase











Today's Challenge – Analyze Some Customer Data

Initial situation:

- Customer data in two datasets:
 - Phone usage
 - Contract information
- Column "Phone" is in both datasets
- Column "churn" encodes whether a customer is happy

Goal:

- Find rules that describe happy and unhappy customers by
 - training a decision tree model
 - calculating aggregations (optional)



Classification

Predict nominal outcomes on existing data (supervised)

Applications

- Churn analysis (yes/no)
- Chemical activity (active/inactive)
- Spam detection (spam/not spam)
- Optical character recognition (A-Z)

Methods

- Decision Trees
- Neural Networks
- Naïve Bayes
- Logistic Regression





lass counts for Target		
Class:	0	1
Count:	5906	5645
otal count: 11551		
hreshold to used for zero probabiliti	es: 0.0	
aussian distribution for Age per clas	s value	
	0	1
Count:	5906	5645
Mean:	49.68557	46.82604
Std. Deviation:	12.27388	10.16363
Rate:	51%	49%
aussian distribution for Available40	1K per class value	
	0	1
Count:	5906	5645
Mean:	0.68134	0.68485
Std. Deviation:	0.466	0.46462
Rate:	51%	49%



Data Mining: Process Overview





Partitioning

- Use to split data into training and evaluation sets
 - Partition by count (e.g. 10 rows) or fraction (e.g. 10%)
 - Sample by a variety of methods; random, linear, stratified

Absolute	100 0
O Relative[%]	50 🗘
• Stratified sampling	S Target
Use random seed	123



Table "d	lefault" – Rows	5775	Spec	- Columns:	13
Row ID	S Marita	S Gender	Estim	Numb	Age
Row0	м	м	90000	0	44
Row7	M	м	60000	2	46
Row9	S	м	70000	1	46
Row10	S	F	70000	1	46
Row13	M	м	100000	3	42
Row14	S	F	100000	3	42
Row15	S	F	30000	1	31
Row17	S	F	20000	2	66
Row18	S	M	30000	2	66
Row20	S	M	40000	2	32
n 7 1	<i>c</i>	-	*****		
👂 🌒 Sei	cond partition (remaining ro	ows) - 0:249	- Partitionin	g
Ser Table "d	cond partition (efault" – Rows	remaining ro	ows) - 0:249 Spec	- Partitionin - Columns:	g 13
Table "d	cond partition (efault" - Rows S Marita	remaining ro 5776	ows) - 0:249 Spec	- Partitionin - Columns:	g 13
Table "d Row ID Row1	cond partition (efault" – Rows S Marita	remaining ro : 5776 S Gender M	Spec 60000	- Partitionin - Columns:	g 13 Age 45
Table "d Row ID Row1 Row2	cond partition (efault" – Rows S Marita S M	remaining ro : 5776 S Gender M M	Spec 50000 60000	- Partitionin - Columns: Numb 1	g 13 45 45
Table "d Row ID Row1 Row2 Row3	efault" – Rows S Marita S M S	remaining ro : 5776 S Gender M M F	Spec 5 Spec 5 Estim 60000 60000 70000	- Partitionin - Columns: I Numb 1 1	g 13 45 45 45 42
Table "d Row ID Row1 Row2 Row3 Row4	efault" – Rows S Marita S M S S S	remaining ro 5776 S Gender M F F	Spec 5pec 5pec 50000 50000 70000 80000	- Partitionin - Columns: Numb 1 1 1 4	g 13 45 45 45 42 42
Table "d Row ID Row1 Row2 Row3 Row4 Row5	efault" – Rows S Marita S M S S S S	remaining ro : 5776 S Gender M M F F F M	Spec 5pec 5pec 50000 50000 70000 80000 70000	- Partitionin - Columns: I Numb 1 1 4 1	9 13 45 45 42 42 42 42
Table "d Row ID Row1 Row2 Row3 Row4 Row5 Row5 Row6	efault" – Rows S Marita S M S S S S S	remaining ro 5776 S Gender M M F F F M F	Spec Spec Estim 60000 60000 70000 80000 70000 70000	- Partitionin - Columns: 1 1 1 4 1 1	9 13 45 45 42 42 42 42 42 45 44
Table "d Row ID Row1 Row2 Row3 Row4 Row5 Row6 Row6 Row8	efault" - Rows s Marita S M S S S S S S	remaining ro 5776 S Gender M F F M F F F	Spec 5pec 5pec 50000 50000 70000 70000 70000 50000 50000	- Partitionin - Columns: 1 1 1 4 1 3	9 13 45 45 42 42 42 42 45 44 46
Table "d Row ID Row1 Row2 Row3 Row4 Row5 Row6 Row8 Row8 Row11	efault" - Rows efault" - Rows S Marita S M S S S S S M	remaining ro 5776 S Gender M F F M F F M M	Spec Spec Estim 60000 70000 80000 70000 60000 60000	- Partitionin - Columns: I Numb 1 1 4 1 1 3 4	9 13 45 45 42 42 42 42 45 44 46 46
Table "d Row ID Row1 Row2 Row3 Row4 Row5 Row6 Row8 Row11 Row12	efault" - Rows S Marita S S S S S S S M M	remaining ro 5776 S Gender M F F M F F M F F F F F F	Spec Spec Spec Spec Spec Spec Spec Source Source Source Source Spec	- Partitionin - Columns: 1 1 1 4 1 1 3 4 2	g 13 45 45 42 42 42 45 44 46 46 46 42
Table "d Row ID Row1 Row2 Row3 Row4 Row5 Row6 Row6 Row8 Row11 Row12 Row12 Row16	efault" – Rows s Marita S M S S S S S M M M M	remaining ro 5776 S Gender M F F F F F F M F F M M	Spec Spec Estim 60000 60000 70000 70000 60000 60000 60000 100000 30000	- Partitionin - Columns: I Numb 1 4 1 3 4 2 1	g 13 45 45 42 42 42 45 44 46 46 46 42 31
Table "d Row ID Row I Row3 Row3 Row3 Row4 Row5 Row6 Row8 Row8 Row11 Row12 Row16 Row19	efault" - Rows efault" - Rows Marita S M S S S S S M M M M S	remaining ro 5776 S Gender M M F F F M F M M M M	Spec	- Partitionin - Columns: Numb 1 1 1 4 1 3 4 2 1 2	g 13 45 45 42 42 42 45 44 46 46 46 42 31 32



Learner-Predictor Motif

- Most data mining approaches in KNIME use a Learner-predictor motif.
- The Learner node trains the model with its input data.
- The Predictor node applies the model to a different subset of data.





Goal: A Decision Tree

Outlook	Wind	Temp	Storage	Sailing
sunny	3	30	yes	yes
sunny	3	25	yes	no
rain	12	15	yes	yes
overcast	15	2	no	no
rain	16	25	yes	yes
sunny	14	18	yes	yes
rain	3	5	no	no
sunny	9	20	yes	yes
overcast	14	5	no	no
sunny	1	7	no	no
rain	4	25	yes	no
rain	14	24	yes	yes
sunny	11	20	yes	yes
sunny	2	18	yes	no
overcast	8	22	yes	yes
overcast	13	24	yes	yes





Decision Tree Learner

Decision Tree Learner

► ਨਾ

options	PMMLSettings Flow Variables
General	
C	ass column S Target 📀
	Quality measure Gini index ᅌ
I	runing method No pruning ᅌ
	Reduced Error Pruning
Min num	per records per node 2 🗘
Number rec	ords to store for view 10,000
	Average split point
	Number threads 8
🗹 Skip n	ominal columns without domain information
Root split	
	Force root split column
Root spli	t column T WebActivity
Binary nomin	al splits
	Binary nominal splits
	Mar dia minut 10 0
	Max #riominal 10



Applying the Model – What are the Outputs?



e Hilite	Navigation	View			
Tal	ole "default" – Re	ows: 879	Spec – Columns: 82	Properties I	Flow Variables
Row ID	I SalePr	S rank	D P (rank=Low)	D P (rank=High)	S Prediction (rank)
10	189000	Low	0.889	0.111	Low
11	175900	Low	1	0	Low
13	180400	Low	1	0	Low
15	212000	Low	0.946	0.054	Low
21	190000	High	0	1	High
22	170000	High	0.2	0.8	High
27	126000	Low	1	0	Low
28	115000	Low	1	0	Low
33	127500	Low	1	0	Low



Decision Tree Predictor

Decision Tree

Predictor

- Takes a decision tree model and apply it to new data
- Check the box to append class probabilities



Evaluation Metrics

- Why evaluation metrics?
 - Quantify the power of the model as a classifier/predictor
 - Compare model configurations and/or models, and select the best performing one
 - Obtain the expected performance of the model for new data
- Different model evaluation techniques are available for
 - Classification/regression models
 - Imbalanced/balanced target class distributions





Overall Accuracy = $\frac{\# Correct Classifications}{\# All Events}$

- The proportion of correct classifications
- Downsides:
 - Only considers the performance in general and not for the different target classes
 - Therefore, not informative when the target class distribution is unbalanced



Confusion Matrix

Arbitrarily define one target class as POSITIVE and the remaining class(es) as NEGATIVE

	Predicted class POSITIVE	Predicted class NEGATIVE
Actual class POSITIVE	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
Actual class NEGATIVE	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

TRUE POSITIVE (**TP**): Actual and predicted class is positive

TRUE NEGATIVE (**TN**): Actual and predicted class is negative

FALSE NEGATIVE (**FN**): Actual class is positive and predicted negative

FALSE POSITIVE (**FP**): Actual class is negative and predicted positive

Use these four statistics to calculate other evaluation metrics, such as overall accuracy, true positive rate, and false positive rate



Scorer

Compare predicted results to known truth in order to evaluate model quality

Scorer	Flow Variables	Job Manager Selection
First Colum	n	
	S Target	٥
Second Col	umn	
	S Prediction (T	arget) ᅌ
Sorting of v	alues in tables	
Sorting stra	tegy: Insertion or	der ᅌ 🗌 Reverse ord
Provide sco	res as flow variables	
	Use name prefix	
Missing val	ues	
	In case of missing v	alues 💽 Ignore
		🔵 Fail

Scorer

▶ 🏠



Scorer

Confusion matrix shows the distribution of model errors

• • •	Confusion Matrix - 0:297 - Scorer				
File Hilite					
Target \ Prediction (Target) 1 0	1 2073 759	0 750 2193			
Correct classifie	ed: 4,266		Wrong classified: 1,509		
Accuracy: 73	3.87 %		Error: 26.13 %		
Cohen's kappa	(к) 0.477				

An accuracy statistics table provides a detailed analysis of model quality

D Accuracy	D Cohen's kappa
?	?
?	?
0.739	0.477
Ī	



Exercise: 13_Training_a_Churn_Prediction_Model

- Read the CallsData.xls and ContractData.csv files
- Join the two data tables based on the columns "Area Code" and "Phone"
- Change the data type of the columns "Area Code" and "Churn" to string
- Partition the data into a training set and a test set
- Train a decision tree to detect customers that are likely to churn
- Apply the mode to the test set and evaluate the model performance



Confirmation of Attendance and Survey

 If you would like to get a "Confirmation of Attendance" please click on the link below*

Confirmation of Attendance and Survey

 The link also takes you to our course feedback survey. Filling it in is optional but highly appreciated!

Thank you!

*Please send your request within the next 3 days

L2-DW Online Course Feedback Confirmation of Attendance Request a confirmation of attendance by filling out this section: Enter name to appear on "Confirmation of Attendance" Your answer Enter email to receive "Confirmation of Attendance" Your answer	Open for Innovation
Confirmation of Attendance Request a confirmation of attendance by filling out this section: Enter name to appear on "Confirmation of Attendance" Your answer Enter email to receive "Confirmation of Attendance" Your answer	L2-DW Online Course Feedback
Request a confirmation of attendance by filling out this section: Enter name to appear on "Confirmation of Attendance" Your answer Your answer Your answer	Confirmation of Attendance
Enter name to appear on "Confirmation of Attendance" Your answer Enter email to receive "Confirmation of Attendance" Your answer	Request a confirmation of attendance by filling out this section:
Your answer Enter email to receive "Confirmation of Attendance" Your answer	Enter name to appear on "Confirmation of Attendance"
Enter email to receive "Confirmation of Attendance" Your answer	Your answer
Your answer	Enter email to receive "Confirmation of Attendance"
	Your answer
Back	Back



Course books are available on **KNIME Press**

https://www.knime.com/knimepress

Download your free copy of the KNIME Beginner's Luck by using the code: **SUMMIT-BOOKS-0423**

https://www.knime.com/knimepress/knime-advanced-luck





KNIME Certification Program

Measure and certify your KNIME and data science expertise https://www.knime.com/certification-program

Four levels L1 – L4 of on-demand certification examinations

KNIME Certifie	Certified	Certified	Open for Innovation Certified
Basic Proficiency in KNIME Analytics Platform	Advanced Proficiency in KNIME Analytics Platform	Proficiency in KNIME Software for Collaboration and Productionizing of Data Science	Text Processing in KNIME Analytics Platform
			L4 🗩 🕄 🔩

Free certification exams for a week (until April 24, 2023) with the codes: L2= SUMMIT-CERT-L2-0423 L3= SUMMIT-CERT-L3-0423 L4= SUMMIT-CERT-L4-0423 (L1 is free)



Stay Connected with KNIME



Blog: knime.com/blog



KNIME Self-Paced Courses: knime.com/knime-self-paced-c ourses



Forum: forum.knime.com



Email: education@knime.com Follow us on social media:





137



KNIME



© 2023 KNIME AG. All rights reserved.



Medium Journal: medium.com/low-code-for-adva nced-data-science

Thank you! education@knime.com



Attachment: How to use a local update site to install extensions

- Download the update site as zip
 - KNIME update site as zip
 - Previous versions of the KNIME update site as zip
 - <u>Community update</u> sites as zip



Open KNIME Analytics Platform and go to the preference page by clicking on

File -> Preferences







- Search for update (upper left search bar) and go to Available Software sites.
- 2. Uncheck all existing software sites.
- 3. Click on Add.. on the upper right.



- 1. Define a name
- 2. Click on Archive and select the folder you've just downloaded
- 3. Click OK
- 4. Click Apply and Close

update		8	Available Software Sites			$\langle \neg \neg \rangle$
▼Install/Update Automatic Updates Available Software Sites		s Sites	type filter text		8	
			Name	^	Locat	Add
			KNIME Analytics Platform 4 KNIME Analytics Platform 4 KNIME Community Extension KNIME Community Extension	1.1 Update Site 1.1 Update Site ons (Experi ons (Trusted)	http:/, https: http:/, http:/,	Edit
			Add Site	ono (n'aotoa)		Remove
	00		Add Site			eload
	Name:	Lokal up	odate Site		Local	nable
	Location:	http://			Archive	
			Car	ncel	OK	.port
				Cancel	Ар	ply and Close

