



Introduzione al software KNIME[®] Analytics Platform



Copyright© 2025 Alfredo Roccato. Tutti i diritti riservati.

I testi, le immagini e la grafica qui presenti sono protetti ai sensi delle normative vigenti sul diritto d'autore, sui brevetti e sulla proprietà intellettuale. È vietata la riproduzione anche parziale e con qualsiasi mezzo senza l'autorizzazione scritta dell'autore.

Per informazioni sui permessi per riprodurre parti del presente lavoro, inviare un messaggio e-mail ad Alfredo Roccato all'indirizzo <u>alfredo.roccato(at)fastwebnet.it</u>. Si prega di indicare quali pagine si desidera utilizzare e per quale scopo.

Questo libro è stato aggiornato per il software KNIME[®] Analytics Platform (Versione 4.5.2 e superiori), R (Versione 4.2.0 e superiori), Python (Versione 3.8.8 e superiori).

Programma

Introduzione

- L'ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice





KNIME (Konstanz Information Miner) è una piattaforma **open-source** sviluppata da Knime AG (<u>http://www.knime.org</u>) per l'analisi dei dati attraverso un'interfaccia grafica che permette all'utente di compiere l'intero processo di analisi, dall'accesso ai dati all'applicazione dei risultati.

L'interfaccia grafica, attraverso un diagramma di flusso, permette l'assemblaggio dei nodi per il **pre-processing dei dati, l'analisi, il modeling, la visualizzazione** dei risultati e il reporting.

KNIME[®] Analytics Platform supporta una vasta gamma di strumenti analitici da quelli **statistici** e di **machine learning** al **text processing**.

Introduzione



La piattaforma applicativa OPEN-SOURCE



Programma

Introduzione

- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Data Mining e Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice





L'Interfaccia grafica



II Workspace

È il luogo fisico (root directory) dove stanno i diagrammi di flusso (Workflow)







II Workspace

Gli **EXAMPLES** sul server remoto sono documentati nella guida online: <u>https://www.knime.com/nodeguide</u>



II Workspace

Dalla versione 4.0 è stata aggiunta nel Workspace la possibilità di **trovare e condividere workflow, nodi e soluzioni applicative**.



Facendo doppio clic su *Private o Public* si apre una pagina dove si può accedere a queste funzionalità.



II Workspace

Ci possono essere diversi Workspace. KNIME, alla partenza, chiede quale aprire. Un Workspace può essere cambiato all'interno di una sessione (KNIME viene chiuso e riaperto)



Il Workflow

KNIME utilizza la programmazione visuale.

Questo comporta che ogni passo di analisi viene rappresentato in un **editor grafico** attraverso un'icona chiamata **Nodo**.

Il Workflow è una sequenza di nodi connessi che vengono aggiunti "trascinandoli" dalla sezione dei nodi e "rilasciati" sulla finestra dell'editor.



Il Workflow

Il Workflow, una volta salvato, compare nel KNIME Explorer.



I Nodi

I Nodi sono connessi tra di loro attraverso le loro porte di input e di output: basta solo 'cliccare' sulla porta di uscita del nodo precedente e rilasciare sulla porta di ingresso di quello successivo.

Node 15

Node 16

I nodi hanno 3 possibili stati rappresentati da colori semaforici:

- Rosso: Presenza di errori in esecuzione o non ancora configurato
- Giallo: Configurato ma non eseguito
- Verde: Eseguito con successo

Ogni nodo può essere configurato ed eseguito con gli appositi comandi ottenuti con il tasto destro del mouse.



Node 17



I Nodi

La configurazione avviene per mezzo di apposite schede (3 o più):

•	La prima e, se presenti, altre	
	schede contenenti le impostazioni	△ Dialog - 0:26 - Column Filter - □ ×
	per l'esecuzione del nodo	File
		Column Filter Flow Variables Memory Policy
		Evelude Select Indude
•	Quella per la gestione delle Flow	Column(s): Search Select all search hits Column(s): Column(s): Search Search
	Variables	S Empliame add all >> S DeptID Salary
•	Quella per la gestione della memoria (vedi appendice)	EmpName << remove all

I Nodi

I Nodi possono avere molte porte.

- Porte di Input e di Output dei dati
 Servono al passaggio dei dati
- Porte di Input e di Output del modello
 Servono al passaggio dei coefficienti e delle statistiche dei modelli
- Porte delle Flow Variables
 Servono per rendere parametrici i Workflow, i nodi vengono impostati dinamicamente.
- Porte delle connessioni ai Database
 Servono per l'accesso di database esterni.
- Porte di output delle immagini







Le estensioni

Le estensioni open-source forniscono funzionalità aggiuntive (attraverso nodi specifici) come l'accesso e l'elaborazione di dati complessi, nonché l'aggiunta di algoritmi avanzati di apprendimento automatico.



Metanodi

Raggruppamento di nodi (Collapse into a Metanode)

 Si possono raggruppare insieme dei nodi all'interno di un Metanodo una volta selezionati con il bottone destro del mouse scegliendo l'opzione *Collapse into Meta Node*.









Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice



Tabella in formato KNIME

Nodo **Table Creator Crea** una tabella in formato KNIME definendo le colonne e inserendo direttamente i valori.

Nodo **Table Reader Legge** una tabella in formato KNIME.



Tabella in formato KNIME

Tipologie di dati più comuni.

- Integer/Double
 - ✓I Interi
 - ✓ D Doppia precisione
- String
 ✓S
 - Testo
- Date/Time
 ✓ jii Data
 ✓ jii Data/Orario

N.B. I dati mancanti vengono rappresentati con il punto interrogativo "?".



File di testo generici

Nodo File Reader

							Perc	corso del fi	le	
Settings Flow	Variables Job	Manager Selecti	on Memory Po	licy						
Enter ASCII da	ata file location:	(press 'Enter' to	update previev	v)					^	
knime://knime	e.mountpoint/Int	roduzione/Appli	cazioni%20-%2	0ContactLab/C	hurn%20Analys	sis%20-%20Co	ntactLab, 🗸	Browse	• •=?	
Basic Settings	Prima c	olonna co	me Row II	D gs for ne	ew location F	Rescan				
read row I	Ds		Column delim	iter: ,	~		- [[Delimitato	re	
read colum	nn headers		🗸 ignore sp	aces and tabs						
			Java-styl	e comments		Single line	comment:			
									New settings for column 'O	×
	Nomi o	lelle colon	ine nella p	rima riga					Column Properties	_
Preview									DON'T include column in output ta	ble
		Click column ł	neader to chang	e column prope	erties (* = name	e/type user set	tings)		Name: Open_Rate	:
Row ID	↓ giorni	send_o	send_o	open_u	🕂 dick_u	open_u	. 🔒 dick_u	D Open_Rate	Type: Number (double)	\sim
Row0	86	21	7	16	0	6	6	0.762	miss. value pattern:	٦
Row1	70	28	7	5	0	6	6	0.179	Format	5
Row2	70	27	6	2	0	5	5	0.074		~
Row3	3	0	1	0	0	0	Δttribut			
Row4	1	0	1	0	0	0	Attribut		Domain	
Row5	1	0	1	0	0	0	0	?		-
									OK Cance	

File di testo delimitati

	Input location
Nodo CSV Reader	Read from Mountpoint V LOCAL V
Legge file esterni in formato csv	Mode File Files in folder Sta in una sottocartella del Workspace
Percorso del file	File //Corso_01_Introduzione/Prove/class.csv
Carattere di delimitazione delle colonne	Reader options Format Autodetect format
Delimitatore di riga	; Column delimiter \r\n Row delimiter " Quote char " Quote escape char
Nomi delle colonne nella prima riga	# Comment char
Prima colonna come Row ID	Image: Support short data rows Image: Prepend file index to row ID
	Preview The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.
	Row ID S Nome S Sesso I Età I Altezza Peso Row 0 Alfredo M 14 175 51
	Row1 Alice F 13 144 38



File Excel

	- Input location
Nodo Excel Reader	Read from Relative to Current workflow
Legge file in formato Excel	Mode File Files in folder Sta nella stessa cartella
(.xls, .xlsx)	del workflow
Dereerse del file	Sheet selection
Percorso del file	Select first sheet with data (Foglio 1)
	○ Select sheet with name Foglio1 ∨
	○ Select sheet at index 0 ÷ (Sheet indexes start with 0.)
Nome del foglio	- Column header
	Table contains column names in row number 1 - (Row numbers start with 1. See 'File
	-Row ID
Nomi delle colonne nella riga 1	Generate row IDs Table contains row IDs in column A
	Sheet area
	Read entire data of the sheet O Read only data in columns from A to and
	rows from 1 to . (S
Lettura di un range	Preview File Content
	Preview with current settings
	() The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.
	Row ID Risposta S Preferenza
	Row0 0 Non ascolto a quell'ora Row1 1 Rock



Input Location (CSV e Excel) – 1/2

Il file sta in un percorso preciso, è **dipendente** dal sistema operativo <workspace>/folder1/folder2/file.ext

Il file sta in una sotto-cartella dell'**attuale Workspace**

<workspace>/folder1/folder2/file.ext

-Input locat	ion
Read from	Local File System 🗸 🗸
Mode	File Files in folder
File	C:\Applicazioni\Knime\Esempi\Dati\Iris.xlsx

-Input locat	on		
Read from	Mountpoint \lor	LOCAL 🗸	>
Mode	● File ○ Files in folder	knime-temp-space My-KNIME-Hub EXAMPLES	
File	/Esempi/Dati/Iris.xlsx	LOCAL	

Il file sta nella stessa cartella dell'**attuale workflow**

<workspace>/cartella1/cartella2/file.ext

-Input locati	on
Read from	Relative to \sim Current workflow \checkmark
Mode	File Files in folder
File	/Iris.xlsx



Input Location (CSV e Excel) – 2/2

Vengono **letti più file** nella stessa cartella

Si possono **selezionare per nomi ed estensioni**

Il file sta in un **URL** (indirizzo di una risorsa su una rete di computer)

-Input locat	ion
Read from	Relative to \checkmark Current workflow \checkmark
Mode	File Files in folder Filter options Include subfolders
Folder	
	(i) Selected 4 of 4 files
-Input locat	on
Read from	Relative to V Current workflow V File sites options V
Mode	○ File ● Files in folder Filter option: □ Case sensitive
Folder	·· File name ·· □ Case sensitive
	(i) Selected 1 of 4 files
-Input locati	on
Read from	Custom/KNIME URL V Timeout 1.000 🜩
	1. Custom/KNIME URL does not support listing/browsing of files. Y
Mode	File Files in folder
UR	https://www.ssc.wisc.edu/~hemken/Rworkshops/read/class.csv

Per approfondimenti: https://docs.knime.com/2021-06/analytics_platform_file_handling_guide/index.html#introduction



Database Relazionali – 1/3





Database Relazionali – 2/3 Dialog - 3:819 - Microsoft Access Connector File Connection Settings JDBC Parameters Advanced Input Type Mapping Output Type Mapping Flow Variables Ci sono altri nodi specifici per la Configuration connessione al database che si Database Dialect: Microsoft Access Driver for Microsoft Access v. 5.0.1 [ID: built-in-msaccess-5.0.1] Driver Name: sta utilizzando. Location Bisogna solo fornire l'indirizzo del C:\Applicazioni\KNIME\Corso_02_Business_Analytics_02_Modelli_Predittivi\Dati\Train.mdb Browse... database e le proprie credenziali, se richieste. Connection settings Flow Variables Memory Policy Connection Microsoft Access Hostname Port **DB Query Reader** Connector 3.306 🜩 localhost \sim Nodo Microsoft Access Connector ₿A 2→ Database name test \sim Accesso Lettura Authentication Tabella mdb Use credentials Nodo MySQL Connector Use username & password Username root Password Nodo SQLite Connector Timezone correction No correction (use UTC) Use local timezone O Use selected timezone Europe/Berlin Misc Allow spaces in column names Validate connection on close Retrieve metadata in configure

Cancel

OK

Apply



Database Relazionali – 3/3

Nodo Database Reader

È un nodo generico di accesso e di lettura al database specificato.

Nodo Database Writer

È un nodo generico di accesso e di scrittura al database specificato.

org.sglite.JI	DBC		
(Additional Databa	se Drivers can be loaded in the k	(NIME preference page.)	
Database URL			
knime://knime	e.workflow/db.sqlite		
Workflow Credent	tials		
User Name			
Password			
TimeZone			
O No Correction	(use UTC) (Use local TimeZ	one	
0.7			
O Timezone:	Europe/Berlin	~	
Misc			
Allow spaces	in column names		
Database Browse	r		SOL Statement
	Fetch Metadata		SELECT * FROM data
A			
Database Reader	GroupBy		
Database Reader	GroupBy		
Database Reader	GroupBy ► <mark>→</mark> ►		
Database Reader	GroupBy	Database Write	91
Database Reader	GroupBy	Database Write	91
Database Reader	GroupBy	Database Write	ər
Database Reader	GroupBy	Database Write	91
Database Reader	GroupBy	Database Write	er



DataBase Relazionali – 3/3 (cont.)

Nodo Database Reader/Writer

Se già presente un nodo dedicato, prende da questo le informazioni per la connessione.

 Database Brow 	ser	SQL Statement
	Fetch Metadata	SELECT * FROM data
- Flow Variable Li	st	
& knime.works	pace	^
S knime.works	pace	^
abase Reader	GroupBy	
	<mark>→</mark> ┣	
		Database Writer

Accesso scrittura

DBMS/tabella

raggruppamento





¹ Necessitano di licenza commerciale.

 \checkmark

 \checkmark

Esportazione

Nodo Table Writer

Scrive una tabella in formato KNIME.

Nodo CSV Writer

Crea un file esterno di tipo tabellare dove le colonne sono separate da un delimitatore.

Nodo Excel Writer

Può essere scritto sia nel formato *Excel 2003* (.xls), sia nel formato *Excel 2007* (.xlsx).

Nodo PMML

Crea un file in formato PMML (Predictive Model Markup Language) per la condivisione di modelli con altri software.





Ordinamento (Sort)

Nodo Sorter

Ordinamento per una o più colonne



Sort by:		
	S -ROWKEY - V	Ascending
Next by:		
	? - DO NOT SORT - 🗸	Ascending
Next by:		
	? - DO NOT SORT - 🗸	Ascending
	Add columns	
	1 _ new columns	
	Sort in memory	
	Move Missing Cells to end of sorte	ed list

Unione

Nodo Concatenate

Aggiunge le righe di una seconda tabella alla prima (tutte le colonne o solo quelle comuni)



Duplicate row ID handling	
◯ Skip Rows	
Append Suffix: _dup	
O Fail Execution	
Column handling	
○ Use intersection of columns	
• Use union of columns	

Join columns

Match () all of the following () any of the following

Top Input ('left' table)



Bottom Input ('rig

Combinazione di 2 tabelle – 1/3

Nodo Joiner

• Le colonne da utilizzare per la combinazione


Accesso ai dati



Combinazione di 2 tabelle – 2/3

Le modalità

		- Include in output	
		Matching rows	Inner join
		Left unmatched rows	\bigcirc
•	Inner ioin	Right unmatched rows	
		L	
		- Indude in output	
		Matching rows	Left outer join
		Left unmatched rows	\bigcirc
•	Left ioin	Right unmatched rows	
	,		
		- Include in output	
		Matching rows	Right outer join
		Left unmatched rows	
•	Right ioin	Right unmatched rows	
		-Include in output	
		Matching rows	Full outer join
		Left unmatched rows	\bigcirc
•	Full ioin	Right unmatched rows	
	i un jonn		

S Prodotto	S Previsione	S Vendite
Sedie	1000	300
Divani	5000	800

S Prodotto	S Previsione	S Vendite
Sedie	1000	300
Divani	5000	800
Poltrone	2000	?

S Prodotto	S Previsione	S Vendite
Sedie	1000	300
Divani	5000	800
?	?	50

S Prodotto	S Previsione	S Vendite
Sedie	1000	300
Poltrone	2000	?
Divani	5000	800
Altro	?	50

Accesso ai dati



Combinazione di 2 tabelle – 3/3

Le colonne da ter

I I /	Joiner Settings Column Selection Performance Flow Variables Memory Policy						
e colonne da tenere	Top Input (left table)						
	Manual Selection Wildcard/Regex Selection Type Selection	Manual Selection Wildcard/Regex Selection Type Selection					
	r Exclude						
	T Filter						
	No columns in this list						
	S Previsione						
	»						
	<						
	Enforce exclusion O Enforce indusion						
	Bottom Input (right table)						
	Titer						
	S Prodotto						
	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>						
	Enforce exclusion O Enforce indusion						
]							
Suffisso se le colonne							
hanno gli stessi nomi	Append custom suffix (right)						
	Chippens cassin series (right)						

Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati

Trasformazione

- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice





Selezione righe (semplice) – 1/3

Nodo Row Filter

 Seleziona in base a una particolare sequenza di valori (pattern matching)

Row0 98357 SALES JOHN DOE 2-JAN-2000 JOHN.DOE@ACME.COM 435-324-3455 250 Row1 98360 HR RIOGER BECK 3-JUL-2001 ROGER.BECK@ACME 435-324-3400 200 Row2 98365 HR JIMMY WONG 4-MAR-2001 JIMMY.WONG@ACME 435-324-3900 300 Row3 98357 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500 Row3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500 Row4 98360 NR GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500 Pile File George Michael@A 435-324-9232 500 500 Indude rows by number Column value matching George Michael@A 44040 44040 Indude rows by number Indude rows by row ID Indude rows by row ID Indude rows by row ID Indude rows by row	Row ID	S EMPNO	S DEPT	S ENAME	S LAST_U	S EMAIL	S TELEPH	SALAR
Row1 98360 HR ROGER BECK 3-JUL-2001 ROGER.BECK@ACME 435-324-3600 200 Row2 98355 HR JIMMY WONG 4-MAR-2001 JIMMY.WONG@ACME 435-324-2390 300 Row3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500 Row3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500 Row3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500 Image: Column to the set: Some matching Image: Column value Image: Column value Image: Column value matching Image: Column value Image: Colu	Row0	98357	SALES	JOHN DOE	2-JAN-2000	JOHN.DOE@ACME.COM	435-324-3455	250
tow2 98365 HR JIMMY WONG 4-MAR-2001 JIMMY, WONG@ACME 435-324-2390 300 tow3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500	low1	98360	HR	ROGER BECK	3-JUL-2001	ROGER.BECK@ACME	435-324-3600	200
ww3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A 435-324-9232 500	ow2	98365	HR	JIMMY WONG	4-MAR-2001	JIMMY.WONG@ACME	. 435-324-2390	300
Dialog - 0:34 - Row Filter File File Column value matching Column to test: © DEPT Include rows by attribute value Discuss by attribute value Exclude rows by attribute value Include rows by number Exclude rows by row ID Watching criteria Include rows by row ID Our separate checking Include rows by row ID Include rows by row ID Our y missing values match	low3	98370	SALES	GEORGE MICHAEL	6-FEB-2002	GEORGE.MICHAEL@A	435-324-9232	500
 filter based on collection elements Matching criteria use pattern matching SAL* case sensitive match contains wild cards regular expression Include rows by number Include rows by number Include rows by number Include rows by row ID Exclude rows by row ID Exclude rows by row ID only missing values match 		Dialog - 0:34	4 - Row Filter	Column value matchin Column to test: S	g DEPT	- [×	
only missing values match		Include rows Exclude rows Include rows Exclude rows Exclude rows Include rows Exclude rows	s by attribute value s by attribute value s by number s by number s by row ID s by row ID	Matching criteria Matching criteria Use pattern mat SAL* case sensitiv Use range check lower bound: upper bound:	iccion elements cching ve match 📿 contai regula	ins wild cards ir expression		
				only missing valu	ues match			

Row ID	S EMPNO	S DEPT	S ENAME	S LAST_U	S EMAIL	S TELEPH	SALARY
Row0	98357	SALES	JOHN DOE	2-JAN-2000	JOHN.DOE@ACME.COM	435-324-3455	250
Row3	98370	SALES	GEORGE MI	6-FEB-2002	GEORGE.MICHAEL@A	435-324-9232	500

S EMPNO

Row ID

S DEPT

S ENAME

S EMAIL

S LAST_U...



SALARY

S TELEPH...

ROGER.BECK@ACME.... 435-324-3600 200

JIMMY.WONG@ACME.... 435-324-2390 300

Selezione righe (semplice) – 1/3 (cont.)

Nodo Row Filter

Row0 SALES JOHN.DOE@ACME.COM 435-324-3455 250 98357 JOHN DOE 2-JAN-2000 Row1 98360 HR ROGER BECK 3-JUL-2001 ROGER.BECK@ACME.... 435-324-3600 200 Row2 HR JIMMY WONG 98365 4-MAR-2001 JIMMY.WONG@ACME.... 435-324-2390 300 Row3 98370 SALES GEORGE MICHAEL 6-FEB-2002 GEORGE.MICHAEL@A... 435-324-9232 500 Dialog - 0:34 - Row Filter _ X File Filter Criteria Flow Variables Memory Policy Column value matching Column to test: | SALARY V filter based on collection elements Matching criteria O use pattern matching •=? case sensitive match contains wild cards Include rows by attribute value O Exclude rows by attribute value regular expression Include rows by number (use range ch Exclude rows by number lower bound: 200 O Include rows by row ID bound: 300 UDDE O Exclude rows by row ID O only missing values match S EMPNO S EMAIL | SALARY Row ID S DEPT S ENAME S LAST_U ... S TELEPH ... Row0 98357 SALES JOHN DOE 2-JAN-2000 JOHN.DOE@ACME.COM 435-324-3455 250

ROGER BECK 3-JUL-2001

JIMMY WONG 4-MAR-2001

- Seleziona in base a una particolare sequenza di valori (pattern matching)
- Seleziona per un intervallo di valori (range checking)

98360

98365

HR

HR

Row1

Row2



Selezione righe (avanzata) – 2/3

Nodo Rule-Based Row Filter

- Operatori relazionali
 , < , >= , <= ,
 MATCHES (RegEX),
 LIKE,
 FALSE, MISSING
- Operatori logici
 OR, AND, XOR



Rule Editor Flow Variables	Memory Policy	
Column List	Category Description	
ROWID ROWINDEX ROWCOUNT S EMPNO S DEPT S ENAME S LAST_UPDATE_DATE S EMAIL	All ~ Function ? </td ? <= ?	^
S TELEPHONE I SALARY Flow Variable List	? AND ? ? IN ? ? LIKE ? ? MATCHES ? ? OR ? ? OR ? ? XOR ? FALSE MISSING ? MOT.2	~
s knime.workspace	Expression 1 /// enter ordered set of rules, e.g.: 2 // \$double column name\$ > 5.0 => FALSE 3 // \$string column name\$ LIKE "*blue*" => FALSE 4 // TRUE => TRUE 5 \$EMPNO\$ >= "98350" AND \$EMPNO\$ <= "98360" AND \$SALARY\$ > 200 => TRUE	^
	Include TRUE matches O Exclude TRUE matches	

_[Row ID	S EMPNO	S DEPT	S ENAME	S LAST_U	S EMAIL	S TELEPH	I SALARY
	Row0	98357	SALES	JOHN DOE	2-JAN-2000	JOHN.DOE@ACME.COM	435-324-3455	250



Selezione righe (avanzata) – 3/3

Rule Editor Flow Variables Memory Policy Nodo Rule-Based Row Splitter Category Description Column List ROWID All \sim ~ ROWINDEX ROWCOUNT Function S EMPNO ? < ? S DEPT ? <= ? Operatori relazionali S ENAME ? = ? S LAST_UPDATE_DATE ? > ? S EMAIL ? >= ? >, <, >=, <=, S TELEPHONE ? AND ? I SALARY ? IN ? MATCHES (RegEX), ? LIKE ? ? MATCHES ? ? OR ? LIKE, ? XOR ? FALSE MISSING ? FALSE, MISSING Flow Variable List NOT 2 s knime.workspace Expression 1 // enter ordered set of rules, e.g.: ^ 2 // \$double column name\$ > 5.0 => FALSE 3 // \$string column name\$ LIKE "*blue*" => FALSE Operatori logici 4 // TRUE => TRUE5 \$EMPNO\$ >= "98350" AND \$EMPNO\$ <= "98360" AND \$SALARY\$ > 200 => TRUE OR, AND, XOR Include TRUE matches O Exclude TRUE matches Rule-based Row Splitter S EMPNO | SALARY Row ID S DEPT S ENAME S LAST_U... S EMAIL S TELEPH ... first output table Row0 98357 SALES JOHN DOE 2-JAN-2000 JOHN.DOE@ACME.COM 435-324-3455 250 second output table Row ID S EMPNO S DEPT S ENAME S LAST_U... S EMAIL S TELEPH ... I SALARY Node 27 Row1 98360 HR ROGER BECK 3-JUL-2001 ROGER.BECK@ACME.COM 435-324-3600 200 Row2 98365 HR JIMMY WONG 4-MAR-2001 JIMMY.WONG@ACME.COM 435-324-2390 300 Row3 98370 SALES GEORGE MI... 6-FEB-2002 GEORGE.MICHAEL@ACM... 435-324-9232 500

Selezione colonne

Nodo Column Filter

• Si indicano le colonne da mantenere in uscita

\rm Dialog - 2:5 - Column Filter			_		×
Column Filter Flow Variables Memory Policy					
Manual Selection Exclude Column(s): Select all search hits SENAME LAST_UPDATE_DATE S TELEPHONE	Wildcard/Regex Selection () Select add >> add all >>	Type Selection Include Column(s): Select all search I EMPNO S EMAIL	hits	Search]
Enforce exclusion	<< remove all	Enforce indusion			



Cambiamento del nome e/o del tipo colonna

Nodo Rename

- Cambia il nome della colonna
- Cambia il tipo della colonna

N.B. Non tutti i cambiamenti di tipo colonne sono ammessi!

À Dialog - 0:36 - Column R File	ename	- 🗆 X
Change columns Flow Variab	les Memory Policy CustomerID	Remove
Filter Options None V	Change: ID	S StringValue V
CustomerID ContractID DateReceived	Price	Remove
Price		IntValue LongValue
		C ComplexNumberValue F FuzzyNumberValue
		BoundedValue StringValue
	OK Apply	Cancel



Cambiamento del tipo colonna¹ – 1/2

Nodo Number to String

Converte colonne numero in colonne stringa

Nodo String to Number

Converte colonne stringa in colonne numero

Nodo Double to Int

Converte colonne numero doppia precisione in colonne numero intero

Exclude	Select	Include
Column(s): Search	add >>	Column(s): Search
ContractID Price	add all >>	CustomerID
	<< remove	
	<< remove all	
<u> </u>		Always include all col

¹ Quelli che non sono ammessi nel nodo Rename.



Cambiamento del tipo colonna – 2/2

Nodo String to Date/Time

Converte colonne stringa in colonne data/ora

Qualora non fosse presente, bisogna ٠ digitare il formato esatto della data.

	🛕 Dialog - 2	∧ Dialog - 2:18 - String to Date/Time — □ ×									
	File										
	Onlines of										
lararia1	Options Flo	w Variables Memory Po	licy								
vorano".		Select column:	data_unsubscrib	e ~							
		Replace	selected column								
	New column name data_unsubscribe_time										
		Date format vyyy-MM-dd									
	- Abort execu	Abort execution									
	Abort e	Abort executionafter this number of unresolved rows 100 牵									
	OK	Apply	Cancel	0							
📐 Parsed time - 2	2:18 - String to	o Date/Time		—		\times					
File											
Table "default" - Ro	ws: 25000 Sp	ec - Columns: 3 Pro	perties Flow Va	ariables							
Row ID	id id	data_unsubs	S registra								
Row1	1	13.dic.2014	2014-09-18			^					
Row2	2	15.lug.2014	2014-05-06								
Row3	3	15.lug.2014	2014-05-06								
Row4	4	29.set.2007	2007-09-26								
Row5	5	23.ott.2007	2007-10-22								

¹ Fare attenzione nel fornire il giusto formato data/orario della colonna utilizzata!

File Table *

Creazione colonne con valori precedenti

Nodo Lag column

Serve per costruire una o più colonne prendendo i valori precedenti *(lag)* di una colonna di riferimento in base all'intervallo specificato.



Row ID	🛃 data	Vendite	Shift	Vendite(-1)	Vendite(-2)	Vendite(-3)
Row0	01.gen.2017	10	3	?	?	?
Row1	01.feb.2017	20	3	10	?	?
Row2	01.mar.2018	30	3	20	10	?
Row3	01.apr.2017	40	3	30	20	10
Row4	01.mag.2017	50	3	40	30	20
Row5	01.giu.2017	60	3	50	40	30
Row6	01.lug.2017	70	3	60	50	40
Row7	01.ago.2017	80	3	70	60	50
Row8	01.set.2017	90	3	80	70	60
Row9	01.ott.2017	100	3	90	80	70
Row10	01.nov.2017	110	3	100	90	80
Row11	01.dic.2017	120	3	110	100	90



Creazione/Modifica valori (Table Lookup)

Nodo Cell Replacer

S Name

Barbara

Carol

Henry

James

Jane

Janet

Row ID ed

S Sex

м

F

м

м

Row0

Row1

Row2

Row3

Row4

Row5

Row6

Row7

Row8

Row ID

Row0

Row1

Row2

Row3

Row4

Row5

Row6

Row7

Sostituisce i valori di una colonna con i valori presi da una tabella di lookup.





Creazione/Modifica valori (Regole, Espressioni, Funzioni) – 2/5

Nodo Math Formula

Crea o sostituisce i valori di **una colonna** con espressioni aritmetiche o funzioni.

A lato un esempio di calcolo della media dei 3 test.

Column List	Row ID	S Student	S StudentID	S Section	Test1	Test2	Final	
ROWINDEX	Row0	Capalleti	Student	1	94	91	87	
ROWCOUNT	Row1	Dubose	1252	2	51	65	91	
Test2	Row2	Engles	1167	1	95	97	97	
Final	Row3	Grant	1230	2	63	75	80	
	Row4	Krupski	2527	2	80	76	71	
	Row5	Lundsford	4860	1	92	40	86	
	Row6	McBane	674	1	75	78	72	
Flow Variable List	exp(pow(abs() sgrt(rand/ Expre	x) x, y) () x) 0 sssion est1\$+ \$Test2 (Ampend Col	2\$+\$Final\$)/3					*
	S Student S Studen	ntiD S Sectio	on Test	L Tes	t2 Fir	nai D M	lean_t	

S Student	S StudentID	S Section	Test1	Test2	Final	D Mean_t
Capalleti	545	1	94	91	87	90.667
Dubose	1252	2	51	65	91	69
Engles	1167	1	95	97	97	96.333
Grant	1230	2	63	75	80	72.667
Krupski	2527	2	80	76	71	75.667
Lundsford	4860	1	92	40	86	72.667
McBane	674	1	75	78	72	75



Creazione/Modifica valori (Regole, Espressioni, Funzioni) – 3/5



Creazione/Modifica valori (Regole, Espressioni, Funzioni) – 4/5

Nodo Column Expressions Expressions Error Handling Flow Variables Memory Policy Expression Type Collection Replace Column Output Column if (column("Valutazione") ... String Giudizio \checkmark if (indexOf(column("Test... String Trovato Expression Editor + column + variable + function 1 ۸ 2 if (column("Valutazione") >= 4) Giudizio="POSITIVO" 3 else Giudizio="NEGATIVO" if (indexOf(column("Testo"), "negativ") >= 0) Trovato="negativ" else Trovato=""

¹ Devono, per sintassi, essere conformi a quelli dello **JavaScript**.

Crea o sostituisce i

con espressioni

Si possono usare

come il costrutto

if ... else ...

strutture di controllo¹

valori di più colonne

aritmetiche o funzioni.

Creazione/Modifica valori (Regole, Espressioni, Funzioni) – 5/5

Nodo Java Snippet

Crea o sostituisce i valori di una colonna con espressioni aritmetiche o funzioni.

A lato esempi di espressioni aritmetiche.

Nella codifica bisogna fare uso di campi Java sia per l'input (c_ ...) sia per l'output (out_ ...).





Creazione/Modifica valori (Funzioni sulle stringhe) – 1/2

Nodo String Manipulation

- strip toglie spazi inizio e fine
- replace sostituisce stringhe
- compare confronta stringhe (<,=,>; -1,0,1)
- count conta occorrenze sottostringa
- indexOf puntatore sottostringa¹
- substr estrattore stringa²
- join unisce stringhe
- length lunghezza stringa
- toInt da stringa a numero
- string da numero a stringa
- joinSep unisce stringhe con separatore
- regexMatcher confronta stringhe con RegEx
- regexReplace sostituisce stringhe con RegEx
- upperCase maiuscolo
- lowerCase minuscolo
- capitalize iniziale maiuscola

Column List	Category	Description
ROWID ROWINDEX ROWCOUNT	All	Strips any whitespace characters from the end of given stri Examples:
S Cust_ID S Name S Address	replaceChars(str, chars, replace) replaceChars(str, chars, replace, modifiers) replaceUmlauts(str, omitE) reverse(str) string(x) string(x) strip(str)	<pre>strip(" KNIME ") = "KNIME" strip("KNIME ", " KNIME") = ["KNIME", "KNIME"] strip(null, "', "a ") = [null, "', "a"] * can be any character sequence.</pre>
Flow Variable List	substr(str)	
& knime.workspace	substr(str, start, length)	v
	Expression	
	<pre>strip(\$Name\$)</pre>	

¹ Il valore va da 0 (se il primo carattere) alla lunghezza della stringa meno 1.

² Estrae sottostringhe dal puntatore per n caratteri, Il puntatore del primo carattere ha valore 0.



Creazione/Modifica valori (Funzioni sulle stringhe) – 2/2

Nodo String Manipulation

Esempi

Campo1	Campo2	Funzione	Trasformazione
M.Rossi@dominio.tlc	1	replace(\$campo1\$,"tld","it")	"M.Rossi@dominio.it"
M.Rossi@dominio.it		substr(\$campo1\$,8)	"dominio.it"
M.Rossi@dominio.it		IndexOf(\$campo1\$,"@")	8
Rossi, Mario		substr(\$campo1\$, 0 ,indexOf(\$campo1\$,","))	"Rossi"
Rossi, Mario		substr(\$campo1\$, indexOf(\$campo1\$,",") +2)	"Mario"
Rossi	Mario	joinSep("_",\$campo2\$,\$campo1\$)	"Mario_Rossi"
(+39) 345-1234567		replaceChars(\$campo1\$,"()- ","")	"+393451234567"
39	40	compare(\$campo1\$,\$campo2\$)	-1
40	40	compare(\$campo1\$,\$campo2\$)	0
41	40	compare(\$campo1\$,\$campo2\$)	1
RSSMRA30A01H501I		regexMatcher(\$campo1\$,"[A-Z]{6}[0-9]{2}[A-Z][0-9]{2}[A-Z][0-9]{3}["True"
RSSMRA30A01H501I		toInt(substr(\$campo1\$,6,2))+1900	1930

Trasposizione righe/colonne – 1/3

Nodo Transpose

Scambia le righe con le colonne.

	S Student	S StudentID	S Section	Test1	Test2	Final		
Row0	Capalleti	545	1	94	91	87		
Row1	Dubose	1252	2	51	65	j 91		
Row2	Engles	1167	1	95	97	97		
Row3	Grant	1230	2	63	75	5 80		
Row4	Krupski	2527	2	De	TD	L Teet1	L Test2	L Final
Row5	Lundsford	4860	1	RO		Testi	Testz	Final
Row6	McBane	674	1	Row0	9	4	91	87
				Row1	5	1	65	91
			X	Row2	9	5	97	97
				Row3	6	3	75	80
				Row4	8	0	76	71
				Row5	9	2	40	86
				Row6	7	5	78	72



Table "default" - Ro	Table "default" - Rows: 3 Spec - Columns: 7 Properties Flow Variables								
Row ID	Row0	Row1	Row2	Row3	Row4	Row5	Row6		
Test1	94	51	95	63	80	92	75		
Test2	91	65	97	75	76	40	78		
Final	87	91	97	80	71	86	72		



Trasposizione righe/colonne – 2/3

Nodi Unpivoting

Si indicano la/e colonne (*Value columns*) i cui nomi vengono riportati nella colonna ColumnNames e i loro valori nella colonna ColumnValues, mantenendo la struttura della colonna indicata come "perno" (*Retained column*).

	S Student	S StudentID	S Section	Test1	Test2	Final					
Row0	Capalleti	545	1	94	91	87	1				
Row1	Dubose	1252	2	51	65	91					
Row2	Engles	1167	1	95	97	97	,				
Row3	Grant	1230	2	63	75	80					
Row4	Krupski	2527	2	80	76	71					
Row5	Lundsford	4860	1	92	40		Row ID	S RowIDs	S ColumnNames	ColumnValues	S Student
Row6	McBane	674	1	75	78		Row0	Row0	Test1	94	Capalleti
							Row1	Row0	Test2	91	Capalleti
							Row2	Row0	Final	87	Capalleti
							Row3	Row1	Test1	51	Dubose
						7	Row4	Row1	Test2	65	Dubose
							Row5	Row1	Final	91	Dubose
	Colu	imn Filter	ranspose				Row6	Row2	Test1	95	Engles
	-	. <u>+ </u> + ▶	▶ ⊞ ▶				Row7	Row2	Test2	97	Engles
Table	Creator	•			-		Row8	Row2	Final	97	Engles
				Value co	olumns		Row9	Row3	Test1	63	Grant
	• N	lode 30	Node 29	Value oc			Row10	Row3	Test2	75	Grant
		<hr/>		test1			Row11	Row3	Final	80	Grant
No	de 28	U	npivoting	toot?			Row12	Row4	Test1	80	Krupski
				lesiz			Row13	Row4	Test2	76	Krupski
				final			Row14	Row4	Final	71	Krupski
			Node 31	Retaine	d colum	ns –					
				Stude	ent						



Trasposizione righe/colonne – 3/3

Nodi Pivoting

Si indicano la/e colonne di raggruppamento (*Group column(s)*) e la/e colonne (*Pivot column(s)*) dai valori delle quali vengono create le nuove colonne i cui valori vengono presi da un'altra colonna (*Manual Aggregation*) secondo la/e funzioni di aggregazione scelte come somma, conteggio, media, deviazione standard, min, max, ...

Row ID	S RowIDs	S ColumnNames	ColumnValues	S Student
Row0	Row0	Test1	94	Capalleti
Row1	Row0	Test2	91	Capalleti
Row2	Row0	Final	87	Capalleti
Row3	Row1	Test1	51	Dubose
Row4	Row1	Test2	65	Dubose
Row5	Row1	Final	91	Dubose
Row6	Row2	Test1	95	Engles
Row7	Row2	Test2	97	Engles
Row8	Row2	Final	97	Engles
Row9	Row3	Test1	63	Grant
Row 10	Row3	Test2	75	Grant
Row11	Row3	Final	80	Grant
Row12	Row4	Test1	80	Krupski
Row13	Row4	Test2	76	Krupski
Row14	Row4	Final	71	Krupski



Transpose

Column Filte

Row ID	S ColumnNames	Capalle	Dubose	Engles	Grant+	Krupski	Lundsf	H McBane
Row0	Final	87	91	97	80	71	86	72
Row1	Test1	94	51	95	63	80	92	75
Row2	Test2	91	65	97	75	76	40	78

Group column(s) ColumnNames Pivot columns Student Manual Aggregation ColumnValues (sum)



Raggruppamento/Sommarizzazione – 1/2

Nodo GroupBy

Nella scheda Groups si selezionano le colonne di raggruppamento

Available column(s)	Select	:	Group column(s)	
Column(s):	Search	add >>	Column(s):	Search
S EMPNO S ENAME		add all >>	S DEPT	
S LAST_UPDATE_DATE S EMAIL S TELEPHONE		<< remove		
SALARY		<< remove all		
	SALARY			
Advanced settings			L	
-				



Raggruppamento/Sommarizzazione – 2/2

Nodo GroupBy

Row0

Row1

Row2

Row3

Nella scheda *Manual Aggregation* si scelgono le colonne sulle quali applicare le funzioni di aggregazione (somma, conteggio, media, dev. standard, min, max, ...).



Row ID	S DEPT	D Mean(S	Min*(S	H Max*(S	D Standa
Row0	HR	250	200	300	70.711
Row1	SALES	375	250	500	176.777

Valori mancanti – 1/2

Nodo Missing Value

Con questo nodo è possibile sostituire i valori mancanti (caratteri o numeri) con uno dei metodi disponibili:

- Interpolazione media
- Interpolazione lineare •
- Media mobile •
- Media •
- Mediana •
- Moda ٠
- Massimo •
- Minimo •
- . . .

oppure

eliminare le righe che li

contengono.



String	Do nothing	~
	Linear Interpolation*	~
Number (integer)	Do nothing	^
	Fix Value	
	Linear Interpolation*	
	Maximum	
	Mean	
	Median	
	Minimum	
	Most Frequent Value	~

String	Remove Row*
Number (integer)	Remove Row*





Nodo Missing Value Column Filter

Rimuove le colonne selezionate che contengono valori mancanti oltre una certa percentuale.



Normalizzazione

Nodo Normalizer

Normalizza i dati numerici con le impostazioni:

- Min/Max,
- Z-Score,
- Decimal Scaling



Nodo Denormalizer

Per ottenere i valori originali







Valori Estremi (Outlier) – 1/2

Nodo Numeric Outliers

Per individuare i valori estremi con il Range Interquartile (IQR) e scegliere se:

- sostituirli con valori mancanti
- eliminare le righe che li contengono

Outlier Treatment	
Apply to All outliers	\sim
Treatment option Replace outlier values	
Remove outlier rows	
Replacement strategy Remove non-outlier rows	
	Outlier Treatment Apply to All outliers Treatment option Replace outlier values Replacement strategy Remove outlier rows Remove non-outlier rows Remove non-outlier rows

Normalizer

▶ 👬

Node 10

Numeric Outliers

Node 24

ġ 🖹



Valori Estremi (Outlier) – 2/2

Nodo Normalizer e Column Expression

La flessibilità del nodo **Column Expression** permette di impostare dei propri valori limite in base alla variabile standardizzata Z.

Z-Score		Intervallo di	confidenza	Errori x 1Mio
1,64485	\rightarrow	90%		50.000
1,95996	\rightarrow	95%		25.000
2,5	\rightarrow	98,7581%	(4 Sigma s.t.)	6.210
2,57583	\rightarrow	99%		5.000
2,69795	\rightarrow	99,3023%	(1,5*IQR)	3.488
4,5	\rightarrow	99,9993%	(6 Sigma s.t.)	3,4





Partizionamento – 1/2

Nodo Auto-Binner

I dati vengono partizionati in bins (gruppi) che possono essere di

ampiezza intervallo $= \frac{(x_{max} - x_{min})}{k}$

uguale ampiezza¹

• uguale frequenza² quantili $n. osservazioni = \frac{N}{k}$

Auto Binner Settings	Number Format Settings	Flow Variables	Memory Policy	
Exclude <i>Filter</i> D sepal_length D sepal_width D petal width		Manual Select	tion () Wildcar	d/Regex Selection Indude Tricer D petal_length
Enforce exclus	ion		>> < <	Enforce inclusion
Fixed number of	fbins			
Number of bins: Equal: Sample quantile: Quantiles (comm	frequency	4, 0.6, 0.8, 1.0	>	
Bin Naming	: Bin 1, Bin 2, Bin 3			



S Petalo_lung [Binned]	Count(Specie)	D Min*(Petalo_lung)	D Max*(Petalo_lung)
Bin 1	37	1.0	1.5
Bin 2	24	1.6	3.9
Bin 3	29	4.0	4.6
Bin 4	30	4.7	5.3
Bin 5	30	5.4	6.9

¹ Sconsigliato in quanto può produrre distribuzioni molto sbilanciate.
 ² Potrebbe non essere possibile ottenere la stessa frequenza a causa di valori ripetuti.



Partizionamento – 2/2

Nodo Numeric-Binner

I dati vengono partizionati in bins (gruppi) che possono essere personalizzati.

Intervals Flow Variables Memory Policy					Numeric E	Binner	GroupBy	
Select Column	petal_length				<mark>,</mark> ≣≣	• • • • •	→ <mark>子</mark> →	
D sepal_length D sepal_width	Add	Remove						
D petal_length (5 bins defined, append new) D petal_width	Bin1 :] -∞ 1,5] Bin2 :] 1,5 4,5]				10	atal las atb)	D Ma S(astal Jacoth)	D Mau & (astal las ath)
pron_most	Bin3 :] 4,5 5,5]		Bin1	37		etal_length)	D Min*(petal_length) 1.0	D Max*(petal_length) 1.5
	Bin4 :] 5,5 6,0] Bin5 :] 6,0 ∞ [Bin2 Bin3	50 38			1.6 4.6	4.5 5.5
			Bin4 Bin5	16 9			5.6 6.1	6.0 6.9
			t.	$\overline{\ }$	\mathcal{I}			
	:] ~	-00 🔺 00 🔺 [🗸						
	Append new column petal_ler	ngth_binned						



Flow Variables (opzionale)

Si presenta talvolta la necessità di voler cambiare dinamicamente le impostazioni dei nodi durante l'esecuzione del Workflow. Questo si può fare attraverso l'utilizzo delle variabili di flusso, le *Flow Variables*.

Queste funzionalità vengono applicate ai nodi con delle porte dedicate e identificate all'interno dei nodi nella relativa scheda.





Flow Variables (opzionale) – 1/2

Nodo Single Selection

In questo esempio vengono impostati, nel nodo Single Selection, i possibili valori che potrà assumere la Flow Variable "lingua". Nel nodo Row Filter, invece che un valore predefinito per selezionare la colonna (column2), si imposta la Flow Variable che lo conterrà.

	Label:	Lingua	Filter Criteria Flow Variables Memor	ry Policy
	Description:	Selezione lingua		Column value matching Column to test: S column2 ~
	Variable Name: Parameter Name: Selection Type: Possible Choices:	lingua single-selection Dropdown ~ EN	 Include rows by attribute value Exclude rows by attribute value 	Inter based on collection elements Matching criteria Matching criteria
Single Selection	-	IT	Include rows by number Exclude rows by number Include rows by row ID	O use range checking
Table Creator Node 22 Row Filter	Default Value:	EN IT	Exclude rows by row ID	O only missing values
Node 20 Node 21		<	1 The "Pattern" parameter is co	ontrolled by a variable.



Flow Variables (opzionale) – 2/2

Nodo Single Selection

Quando viene eseguito il nodo Single Selection, si apre un modulo (quickform) dove si può selezionare il valore desiderato, che sarà quello utilizzato per filtrare le righe della colonna indicata nel nodo Row Filter (column2).



Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice



Evidenziare parti selezionate (Highlighting)

Quando si lavora su dati visualizzati su più finestre, può essere utile poterli selezionare su una vista e vederne i collegamenti su tutte le altre in modo dinamico.

Si selezionano prima dalla tabella o dal grafico i dati da collegare e poi dal menu dei risultati si sceglie *Hilite → Hilite Selected*



ble "default" -	Rows: 19 Spec	c - Columns: 5	Properties Flo	w Variables	
Row ID	S Nome	S Sesso	Età	Altezza	Peso
Row0	Alfredo	м	14	175	51
Row1	Alice	F	13	144	38
Row2	Barbara	F	13	166	44
Row3	Carolina	F	14	160	46
Row4	Enrico	м	14	161	46
Row5	Giacomo	м	12	146	38
Row6	Gianna	F	12	152	38
Row7	Chiara	F	15	159	51
Row8	Davide	м	13	159	38
Row9	Giovanni	м	12	150	45
Row 10	Gabriella	F	11	130	23
Row11	Patrizia	F	14	163	41
Row12	Luisa	F	12	143	35
Row13	Maria	F	15	169	51
Row 14	Filippo	м	16	183	68
Row 15	Roberto	м	12	165	58
Row 16	Ronald	м	15	170	60
Row17	Tommaso	м	11	146	39
Row 18	Marco	M	15	169	51

"M"->Blu, "F"-> Rosa

D,

• 22

Peso * Altezza
Funzionalità

Nodi di visualizzazione





Utility/Diagrammi

Nodo Color Manager

assegna un colore ad ogni valore di colonna

	Color Settings	Flow Variables	Memory Policy			
	- Selectione Colu	umn				^
(S Sesso					
	Nominal				○ Range	
	M					
	F					
					Preview	
						- 1
	Colori cam <u>p</u> ione	HSV HSL R	GB CMYK Alph	na		
				┤┥┥┥┥		-
	<					>

Row ID	S Nome	S Sesso	Età	Altezza	Peso
Row0	Alfred	М	14	175	51
Row1	Alice	F	13	144	38
Row2	Barbara	F	13	166	44
Row3	Carol	F	14	160	46
Row4	Henry	м	14	161	46
Row5	James	м	12	146	38
Row6	Jane	F	12	152	38
Row7	Janet	F	15	159	51
Row8	Jeffrey	м	13	159	38
Row9	John	м	12	150	45
Row 10	Joyce	F	11	130	23
Row11	Judy	F	14	163	41
Row12	Louise	F	12	143	35
Row13	Mary	F	15	169	51
Row14	Philip	м	16	183	68
Row15	Robert	м	12	165	58
Row16	Ronald	м	15	170	60
Row17	Thomas	м	11	146	39
Row 18	William	М	15	169	51

Histogram

27.5

25 30

35 40 45 50 55 60 65 70

Fit to size

Backmand Color

Citize and allasin

0,325 0,300 0,275 0,250

0,225 0,200 0,175 0,150 0,125 0,100 0,075 0,050 0,025 0,000

٠

36.5

45.5

54.5

63.5

Istogrammi

Nodo Histogram

Nodo Histogram Chart (JFreeChart)

Nodo Interactive Histogram (local)



Diagrammi a barre

Nodo Bar Chart



Nodo Bar Chart (JFreeChart)



Nodo GroupBy + GroupBy Bar Chart (JFreeChart)



¹ Solo per colonne di tipo stringa.





¹ Solo per colonne di tipo carattere.



Introduzione al software KNIME Analytics Platform



Nodo Scatter Matrix (local)



Grafici a linee

Nodo Line Plot





Diagrammi a scatola (e baffi)

Nodo Conditional Box Plot



Nodo Box Plot (local)



Utility/Diagrammi

Nodo Image Writer (Port)

esporta il grafico nel formato del nodo che lo precede (PNG o SVG)



Impaginare più grafici

Nodo Component

Con questo nodo è possibile "incapsulare" più grafici insieme e visualizzarli attraverso una **griglia personalizzata**.



Insieme di grafici Diagrammi di vario tipo Visual Layout Basic Layout Advanced Lay ♥ ▣ < 🔐 🖗 🗘 🖸 🚺 t or click Dpens a dialog to edit usage and layout of Compone Welcome to KNIME Analytics Platform ut or click Text Output **Pie/Donut Chart** 0 0 - 🌗 Scatter Plot Component Output Component Inp Titolo • 🗉 Bar Chart 0 8) .ii × Scatter Plot Node 1

Lo si apre e lo si configura

●F ●M









Per approfondimenti:

https://docs.knime.com/2020-07/analytics_platform_components_guide/index.html#widget-nodes

Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice





Statistiche descrittive per raggruppamenti

Nodo GroupBy

Le statistiche descrittive, come già visto in precedenza, si possono ottenere anche per aggregazioni di valori di altre colonne.

Groups Manual Aggregation	Pattern Based Aggregation	Type Based Aggregation	
Groups Manual Aggregation Aggregation settings Available columns Nome Età Altezza Peso	Pattern Based Aggregation Select add >> add all >> < <remove< td=""> <<<remove all<="" td=""></remove></remove<>	Type Based Aggregation To change multiple umn Peso Peso Altezza Altezza Altezza	columns use right mouse click f Aggregation (click to change) Iedian Iean Iedian Iean Itandard deviation
Advanced settings	agregation method (column na	me) V Fnable hilitir	

S Sesso	D Median(Peso)	D Mean(Peso)	D Standard deviation(Peso)	D Median(Altezza)	D Mean(Altezza)	D Standard deviation(Altezza)
F	41	40.778	8.743	159	154	12.806
м	48.5	49.4	10.2	163	162.4	12.456

Calcolo Quantili

Nodo GroupBy

Con i quantili si possono calcolare qualsiasi divisione della popolazione in n parti di uguale frequenza (per esempio i quartili, quintili, decili, percentili, ...). In particolare:

quantile	quartile	quintile	decile	percentile	
0	0	0	0	0	minimo
0,025					
0,10			1	10	
0,20		1	2	20	
0,25	1			25	Q1
0,4		2	4	40	
0,5	2		5	50	mediana
0,6		3	6	60	
0,75	3			75	Q3
0,8		4	8	80	
0,9			9	90	
1	4	5	10	100	massimo

Calcolo Quantili

Nodo GroupBy

Per il calcolo dei quantili ci sono ben 9 metodi¹. Si possono selezionare da *Edit* \rightarrow *Aggregation Settings* \rightarrow *Advanced*.

Groups Manual Aggregation	Pattern Based Aggree	gation Type Based	d Aggregation	Basic Advanced Quantile: 0,25 +	Basic	Advanced
Available columns	Select	To d	hange multiple columns	OK Cancel	enu.	R_2 ^ K C ^{R_3} 2
I Età I Altezza I Peso		I Altezza	Quantile Median Quantile		Edit Edit Edit	R_5 R_6 R_7 R_8 R_9 ↓
	add >>	I Altezza I Peso I Peso	Maximum Quantile Median		Edit Edit	
	add all >>	Peso Peso	Quantile Maximum		Edit	

S Sesso	D Altezza (0.25	D Altezza (Med	D Altezza (0.75	Altezza (Max*)	D Peso (0.25	D Peso (Med	D Peso (0.75	Peso (Max*)
F	144.0	159.0	163.0	169	38.0	41.0	46.0	51
M	152.25	163.0	169.75	183	40.5	48.5	56.25	68

¹ <u>https://en.wikipedia.org/wiki/Quantile#Estimating_quantiles_from_a_sample</u>

Х

Visualizza le statistiche descrittive di una tabella Knime delle colonne d'interesse.

Nodo Data Explorer

Esplorazione – 1/2

S Nome	S Sesso	🖡 Età	Altezza	D Peso
Alfred	М	14	175	51
Alice	F	13	144	38
Barbara	F	13	166	44
Carol	F	14	160	46
Henry	М	14	161	46
James	М	12	146	38
Jane	F	12	152	38
Janet	F	15	159	51
Jeffrey	м	13	159	38
John	М	12	150	45
Joyce	F	11	130	23
Judy	F	14	163	41
Louise	F	12	143	35
Mary	F	15	169	51
Philip	м	16	183	68
Robert	м	12	165	58
Ronald	М	15	170	60
Thomas	м	11	146	39
William	м	15	169	51



Introduzione al software KNIME Analytics Platform



Esplorazione – 2/2

Nodo Data Explorer

Le colonne indicate come escluse non vengono portate in uscita.

S Nome	S Sesso	Età	à	Alte	zza) Pes	D											
Alfred	М	14		175	51													
Alice	F	13		144	38	3												
Barbara	F	\land Data Explore	r View	-														- 0
Carol	F	Numeric	Nominal U	ata Preview														
Henry	м															Sea	rch:	
James	м		Exclude				Standard					Overall	No.	No.	No.	No.	No.	
Jane	F	Column	Column	Maimum	Maximum 1	Mean 1	Deviation	- 41	Variance 1	Skewness	Kurtosis	Sum 1	zeros	missings 1	NaN 1	+== 11	-0 []	Histogram
Janet	F	Et	×	"	10	13.316	1.493		2.228	0.064	-1.111	253	0	0	0	0	0	
Jeffrey	м		\smile															
John	м	Altezza	-	130	183	158.421	13.001		169.035	-0.287	-0.054	3010	0	U	0	0	0	n La
Joyce	F		-			15.010	10.071				0.704							n efti titten
Judy	F	Peso	10	23	68	45.316	10.274		105.561	0.201	0.704	801	0	0	0	0	0	In .
Louise	F																	
Mary	F	15	Numeric	Nominal	Data Preview													
Philip	м	16			\sim													Search:
Robert	м	12	Colur	mn	Lt Exclude Col	umn		No mis	sings	It	Unique values		I†.	All nominal value	•		It Fr	equency Bar Chart
Ronald	м	15	Nome					0	anga	+1	10		+1	Thomas			•1 ••	equency but enant
Thomas	м	11	NOTIR		Ĩ						10			Judy,				
William	м	15			\sim									Alice,				
			Sessi	0	•			0			2			[], Ronald, Alfred, Henry, Jeffrey, Mary M, F				

Row ID	S Sesso	Altezza	Peso
Row0	м	175	51
Row1	F	144	38
Row2	F	166	44
Row3	F	160	46
Row4	м	161	46
Row5	м	146	38
Row6	F	152	38
Row7	F	159	51
Row8	м	159	38
Row9	м	150	45
Row10	F	130	23
Row11	F	163	41
Row12	F	143	35
Row13	F	169	51
Row14	м	183	68
Row15	м	165	58
Row16	м	170	60
Row17	м	146	39
Dow 19	M	160	E1





Statistiche descrittive

Nodo Statistics

Column	Min	Mean	Median	Мах	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No∞	Histogram
Età	11	13,3158	?	16	1,4927	0,0636	-1,1109	0	0	0	
Altezza	130	158,4211	?	183	13,0013	-0,2869	-0,0536	0	0	0	
Peso	23	45,3158	?	68	10,2743	0,2006	0,7035	0	0	0	23 68

Statistiche descrittive

Nodo Crosstab

Cross Tabulation of Sesso by Età

Frequency Column Percent	Adolescenti	Pre-Adol	escenti	Total	Frequency
F	6		3	9	Deviation
	50%	4	2,8571%		Percent
М	6		4	10	Row Percent
	50%	5	7,1429%		Column Percent
Total	12		7	19	Cell Chi-Square
					Max rows:
					10 🖵
					10 -
itatistics for Table (of Sesso by Età				
Statistic		DF	Val	ue	Prob
Chi-Square		1		0,0905	0,7636
Chi-Square Fisher's Exact Test (2-ta	й)	1		0,0905	0,7636 1





Nodo Linear Correlation



Row ID	D Età	D Altezza	D Peso
Età	1	0.811	0.739
Altezza	0.811	1	0.88
Peso	0.739	0.88	1



Correlazione lineare – 2/2

Nodo Correlation Filter

Rimuove la/e colonne ridondanti che sono sopra un valore di soglia di correlazione.



In questo esempio, la colonna *peso* non viene portata in uscita in quanto correlata positivamente con la colonna *altezza* (*r*=0,88) e superiore al valore di soglia.



	Row ID	S Nome	S Sesso	Età	Altezza
ĺ	Row0	Alfred	М	14	175
	Row1	Alice	F	13	144
	Row2	Barbara	F	13	166

Regressione lineare

Nodi Linear Regression Learner e Linear Regression Predictor

Calcola una regressione lineare multipla.



Campionamento

Nodo Row Sampling

Si può fare prendendo n righe o la loro percentuale, in modo casuale o stratificato.

Con il **campionamento stratificato** la distribuzione dei valori della colonna selezionata è (approssimativamente) mantenuta nella tabella di uscita.

Se si vogliono ottenere sempre le stesse tabelle in uscita, si può usare un **seme** casuale.

Sampling Method Flow Variables Memory Policy	
○ Absolute	100 🜩
Relative[%]	10 📥
○ Take from top	
O Linear sampling	
O Draw randomly	
Stratified sampling	S species \vee
Use random seed	12345

Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche

Machine Learning

- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione

Appendice



Partizionamento

Nodo Partitioning



First partition	Flow Variables	Memory Policy
Choose size o	f first partition	
		100 🜩
Relative[9	%]	80 🜩
◯ Take from	1 top	
O Linear san	npling	
O Draw rand	domly	
Stratified	sampling	S Dolore 🗸
Use rando	om seed	12345

Regressione Logistica – 1/3

Nodi Logistic Regression Learner e Logistic Regression Predictor



Esempio: studio farmaco per nevralgia



Settings Advanced Flow Variables Memory Policy

Regressione Logistica – 2/3

Nodi Logistic Regression Learner e Logistic Regression Predictor



S Dol	ore S Terapi	a Sess	D Eta	D P (Dolo	D P (Dolore=Si).	S Pr	ediction (Dolore)
No	1. Placebo	2. F	68	0.411	0.589	Sì	
Sì	3. Farmaco	B 2. F	76	0.914	0.086	No	
No	3. Farmaco	B 2. F	66	0.988	0.012	No	
No	2. Farmaco	A 1. M	62	0.855	0.145	No	
Sì	1. Placebo	1. M	83	0.005	0.995	Sì	
No	2. Farmaco	A 1. M	70	0.539	0.461	No	
Sì	3. Farmaco	B 1. M	77	0.538	0.462	No	
No	3. Farmaco	B 2. F	69	0.978	0.022	No	
Sì	3. Farmaco	B 1. M	75	0.636	0.364	No	
No	3. Farmaco	B 1. M	70	0.828	0.172	No	
Sì	1. Placebo	2. F	72	0.237	0.763	Sì	
No	2. Farmaco	A 2.F	69	0.914	0.086	No	



Regressione Logistica/Matrice di confusione – 3/3

Nodo Scorer (Java Script)



onfusion Matrix								
Rows Number : 24	No (Predicted)	Sì (Predicted)						
No (Actual)	12	2	85.71%					
Sì (Actual)	4	6	60.00%					
	75.00%	75.00%						

Class Statistics

Class	True Positives	False Positives	True Negatives	False Negatives	Precision	Sensitivity	Specificity	F-measure
No	12	4	6	2	75.00%	85.71%	60.00%	80.00%
Sì	6	2	12	4	75.00%	60.00%	85.71%	66.67%

Overall Statistics

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
75.00%	25.00%	0.471	18	6



Albero di decisione – 1/3

Nodi Decision Tree Learner e Decision Tree Predictor

Esempio: giocare a tennis

S Scelta
onGioca
onGioca
ioca
ioca
ioca
onGioca
ioca
onGioca
ioca
onGioca

Excel Reader (XLS)

XLS

Node 45





Albero di decisione – 3/3



S Meteo	S Temper	S Umidità	S Vento	S Scelta	D P (Scelta=NonGio	D V P (Scelta=Gio	S I_Scelta
Coperto	Freddo	Normale	S	Gioca	0	1	Gioca
Coperto	Caldo	Normale	N	Gioca	0	1	Gioca
Pioggia	Tiepido	Alta	N	Gioca	0.333	0.667	Gioca
Sole	Tiepido	Alta	N	NonGioca	0.333	0.667	Gioca
Pioggia	Tiepido	Normale	N	Gioca	0.333	0.667	Gioca
Sole	Tiepido	Normale	S	Gioca	1	0	NonGioca
Pioggia	Tiepido	Alta	S	NonGioca	1	0	NonGioca



Validazione modelli – Curva ROC

Indicando per ogni nodo Predictor un apposito suffisso è possibile confrontare insieme diversi modelli:



Cluster Analysis (gerarchica) – 1/2

Nodo Hierarchical Clustering

Esempio: segmentare clienti banca per prodotti posseduti







Cluster Analysis (gerarchica) – 2/2







- Il Cluster A, "Risparmiatori Tradizionali". È caratterizzato da libretti, obbligazioni e fondi.
- || Cluster B, "Indebitati".

È caratterizzato da *mutui e prestiti*.

- Il Cluster C, "Investitori".

È caratterizzato in gran parte da *fondi e obbligazioni*.
Machine Learning

- Exclude



Cluster Analysis (non gerarchica) – 1/2

Nodo k-Means

Esempio: collocazione di nuove agenzie bancarie

SSLL	t_att_tot	Tasso_AGR	T_IND	T_SERV	Sport. BANC	DEPOSITI/RESIDEN1	IMPONIB/RESIDENT	IMPONIB/CONTR
BUSTO ARSIZIO1	45.9	0.49	55.6	43.9	62	24.0	23.6	26.
GALLARATE	47.5	0.58	60.5	39.0	85	21.1	13.8	24.
LUINO	44.1	2.04	47.9	50.1	14	12.1	5.5	20.
SESTO CALENDE	44.7	1.55	56.2	42.3	42	15.9	6.9	23.
VARESE1	45.4	1.07	49.6	49.4	104	23.7	17.1	27.
MILANO1	45.7	0.65	48.9	50.5	35	23.6	22.9	28.
VARESE2	45.6	1.29	58.8	39.9	4	12.9	5.1	25.
BELLAGIO1	44.0	5.19	41.4	53.4	3	16.5	7.2	20.
CAMPIONE D'ITALIA	43.9	3.97	35.0	61.0	7	17.0	7.8	20.
СОМО	45.9	1.13	49.8	49.0	207	23.0	21.0	26.
MENAGGIO	42.5	4.43	45.9	49.6	18	17.1	5.4	20.
PORLEZZA	42.8	3.78	47.0	49.2	7	15.4	4.7	13.
MORBEGNO1	44.4	8.39	48.5	43.2	1	8.2	3.0	17.
DESIO1	46.6	1.03	60.7	38.3	24	20.8	16.6	23.
BORMIO	42.3	2.11	32.9	65.0	14	28.8	32.1	20.
CHIAVENNA	42.3	4.71	45.8	49.5	11	16.5	8.5	19.
CHIESA IN VALMALEN	41.3	3.38	47.2	49.4	4	14.6	7.6	16.
MORBEGNO2	42.3	4.10	48.5	47.4	23	19.8	15.1	21.
SONDALO	42.7	3.24	38.9	57.8	7	13.4	6.0	21.
SONDRIO	42.7	4.75	32.5	62.7	27	21.7	23.7	23.
TIRANO	43.5	7.32	31.4	61.3	7	21.6	15.2	22.
EDOLO1	44.4	2.51	20.9	76.6	2	25.5	17.2	17.
BUSTO ARSIZIO2	46.3	0.76	56.2	43.0	103	18.9	13.8	25.
LECCO1	44.2	1.57	54.9	43.5	2	12.4	5.6	26.
DESIO2	46.9	0.55	53.6	45.9	172	18.9	13.1	24.
MILANO2	46.5	0.46	34.8	64.7	1499	32.2	65.0	30.
SANTANGELO LODIG	39.7	3.98	34.3	61.8	3	28.3	12.7	23.



5 cluster

Machine Learning



Cluster Analysis (non gerarchica) – 2/2



Variabile	Statistica	TOTALE	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
	Numero	88	12	20	12	28	16
	%	100	13	21	13	31	22
DEPOSITI/RESIDENTI	Media	21,1	22,4	23,0	25,5	16,9	21,9
IMPONIB/CONTRIB	Media	22,0	26,0	22,4	22,4	21,1	20,3
IMPONIB/RESIDENTI	Media	13,4	19,1	17,5	19,5	8,2	10,0
SPORT. BANC	Media	31,7	119,8	25,0	33,9	13,2	10,9
T_IND	Media	48,6	51,3	58,6	33,4	49,2	45,6
T_SERV	Media	44,7	47,2	35,9	61,3	46,3	39,4
TASSO_AGR	Media	6,7	1,5	5,5	5,3	4,5	15,0
T_ATT_TOT	Media	43,9	45,8	45,7	42,8	43,2	42,8

Dal confronto delle medie di ciascun segmento con le rispettive medie della popolazione risulta di particolare interesse il cluster che ha:

- un valor medio degli occupati nell'agricoltura superiore al valor medio della popolazione
- un valor medio del numero di sportelli bancari inferiore al valor medio della popolazione
- un valor medio dei *depositi bancari per abitante superiore* al valor medio della

Associazioni – 1/3

Nodo Association Rule Learner

Esempio: regole di associazioni prodotti del settore alimentari di un supermercato

Cliente	Prodotto
1	[Zucchero,Uova,Latte,Farina,Lievito]
2	[Zucchero,Uova,Olio,Farina,Lievito]
3	[Uova,Olio,Latte,Farina]
4	[Zucchero,Uova,Latte,Lievito]
5	[Zucchero,Uova,Olio,Latte,Farina,Lievito]
6	[Zucchero,Uova,Farina]
7	[Zucchero,Uova,Olio,Farina,Lievito]
8	[Uova,Olio]
9	[Uova,Olio,Latte,Farina,Lievito]
10	[Zucchero,Uova,Latte,Farina]



Conf. >= 10



Associazioni – 2/3

Nodo Association Rule Learner

Regola	Supporto	Confidenza	Lift Le	Lift Leverage	
{Uova} => {Farina}	0,8	0,8	1,0	0	
{Farina} => {Uova}	0,8	1	1,0	0	
{Olio,Farina} => {Lievito}	0,4	0,8	1,3	0,1	
{Zucchero,Uova} => {Farina}	0,6	0,86	1,07	0,04	
{Zucchero,Uova,Latte} => {Lievito) 0,3	0,75	1,25	0,06	

Una regola credibile ha un valore alto di confidenza e di supporto e un valore di Lift maggiore di 1. Le regole che hanno un livello alto di confidenza, ma che hanno scarso supporto, devono essere interpretate con cautela, di nessuna utilità se hanno Lift = 1 o Leverage = 0.

Machine Learning



Associazioni – 3/3

Regole di Associazione - Diagramma dei collegamenti

Nodo Association Rule Learner

Viene anche usato un nodo **R View** per visualizzare graficamente le associazioni dei prodotti

R View (Table)	Image Writer (Port)
→ = R	



Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice





Trasformazione del testo in una matrice numerica

Con Knime si possono fare analisi su dati non strutturati con i nodi di **Text**

Row ID	S Text
Row0	Cordialità del medico e del personale.
Row1	Il servizio presso la clinica oculistica è stato veloce.
Row2	Il medico e le altre persone sono stati molto, molto cordiali.
Row3	Il tempo di attesa è stato eccellente e il personale è stato molto disponibile.
Row4	Il modo in cui è stata fatta la terapia.
Row5	Nessun problema nel prenotare una visita.
Row6	Velocità nel servizio.
Row7	Il modo in cui sono stato trattato e i miei referti.
Row8	Nessun tempo di attesa, i referti sono stati consegnati velocemente, ottima terapia.



Applicazioni



Alla matrice termini/documenti si applicano tecniche di visualizzazione

 Parole rilevanti (Tag Clouds)



problema cordialità prenotare disponibile medico cordiali veloce eccellente servizio terapia modovelocità referti presso coulistica consegnati visita ottima personale trattato velocemente nessun fatta altre clinica



Applicazioni

Alla matrice termini/documenti si applicano modelli di machine learning

Clustering



Row ID	D ▲ duster_0	D 🛦 cluster_1	Conster_2	
nessun referti consegnati velocemente ottima terapia	0	0	0.05	
eccellente personale disponibile	0	0	0.053	
modo fatta terapia	0	0	0.047	
modo trattato referti	0 0		0.047	
nessun problema prenotare visita	0	0	0.055	
cordialità medico personale	0.073	0	0.033	
medico altre cordiali	0.285	0	0	
velocità servizio	0	0.134	0	
servizio presso clinica oculistica veloce	0	0.149	0	

- cluster_0 ("Gentilezza"): cordialità medici e personale
- cluster_1 ("Efficienza"): servizio veloce
- cluster_2 ("Efficacia"): trattamento e risultati

Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice



Integrazione con R e Python



Nodi di Scripting (R)

In Knime si possono integrare funzionalità del software R.

Nodo **R Snippet** Permette l'esecuzione di uno script R da parte di Knime.

In questo esempio si testa la differenza tra due proporzioni (H₀: n2/N2 > n1/N1) usando la funzione '**prop.test**'.



Nodi di Scripting (R)

Nodo **R Learner** Esegue uno script R che fornisce in uscita un modello.

Nodo **R Predictor** Esegue uno script R utilizzando il modello creato dal nodo R Learner per effettuare previsioni.

In questo esempio viene utilizzato un albero di decisione con la funzione rpart del package **'rpart'**.



Integrazione con R e Python

Nodi Scripting (Python)

Nodo **Python Learner** Esegue uno script Python che fornisce in uscita un modello.

Nodo **Python Predictor** Esegue uno script Python utilizzando il modello creato dal nodo Python Learner per effettuare previsioni.

In questo esempio viene utilizzato il package '*scikit-learn*'.



¹ Vedi <u>Appendice</u>.

Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione
- Appendice



Rilascio in produzione

Nodi PMML



Nodo PMML Predictor

Effettua previsioni della classe di appartenenza su un nuovo insieme di dati

Programma

- Introduzione
- Ambiente operativo
- Accesso ai dati
- Trasformazione
- Visualizzazione/Reporting
- Analisi statistiche
- Machine Learning
- Text Processing & Analytics
- Integrazione con R e Python
- Rilascio in produzione

Appendice





Lingua nazionale

Permette di elaborare e visualizzare numeri e date/orari formattati secondo le convenzioni della lingua e dell'area geografica utilizzate.

- Aggiungere nel file Knime.ini la lingua e la regione desiderata
 Ad esempio, per la lingua italiana:
 - -Duser.language=it -Duser.region=IT



Utilizzo della memoria

Opzioni

- Cambiare l'impostazione di -Xmx nel file Knime.ini
 Ad esempio, per portarlo da 1 gigabyte a 4 gigabyte
 - -Xmx1024m → -Xmx4g
- Per elaborazioni onerose dal punto di vista computazionale si può scegliere dalla scheda Memory Policy di configurazione del nodo l'opzione "Keep all in memory".

뇌 Dialog - 0:: ile	26 - Column <mark>F</mark> il	ter
Column Filter	Flow Variables	Memory Policy
Select memo	ory policy for dat	ta outport(s)
Keep al	in memory.	
○ Keep or	nly small tables in	memory.
O Write ta	ables to disc.	

 In questa pagina si trovano ulteriori suggerimenti e consigli per ottimizzare l'esecuzione dei Workflow:

https://www.knime.com/blog/optimizing-knime-workflows-for-performance



Integrazione con R – 1/2

Installazione in R del package Rserve



Integrazione con R - 2/2

Indicare nelle preferenze il percorso dell'installazione di R

Δ	KNIME Analytics Platform							
File	Edit View Help							
2	New	Ctrl+N	A Preferences					×
	Save	Ctrl+S	to us a Ciltara to a t	- II	_	~	-0	
	Save As		type filter text	- 1	R	(• 🖓	• •
R	Save All	Ctrl+Shift+S	> Ant ^	^	KNIME Diperformance			
	Close All	Ctrl+Shift+W	> Data Management		KNIME R preferences			
	Print	Ctrl+P	> General		Path to R Home C:\Program Files\R\R-3.2.2		Browse	e
	Import KNIME Workflow		> Help		Description by the size limit (in MD = 0.6 coupling itself) 255			
	Export KNIME Workflow		> Install/Update		Riserve receiving buffer size limit (in MB 0 for unlimited, 200			
	Switch Workspace	>	> Java					
	Preferences		✓ KNIME					
Ŕ	Export Preferences		Databases					
÷	Import Preferences		KNIME Explorer					
	Install KNIME Extensions		> KNIME GUI					
	Update KNIME		KNIME Store					
	Exit		Master Key					
			Meta Info Preference					
			Palladian Geocoders					
			Palladian Location E					
			Palladian Ranking Se					
			Palladian Web Searc					
			Preferred Renderers					
			R					
			> Textprocessing					
			> Workflow Coach					
			> Plug-in Development					
			> Report Design 🗸	~				
			< >		Restore Def	aults	Apply	у

- Integrazione con Python¹ 1/2
 - 1. Installare il package Anaconda: <u>https://www.anaconda.com/download/</u>
 - Scaricare questo <u>file di configurazione</u> (Click tasto destro → Salva link con nome...) in una cartella (p.e in *C:\Temp*)
 - 3. Aprire il prompt di Anaconda ed eseguire questo comando: conda env create -f C:\Temp\py36_knime.yml
 - 4. Creare il file **py36.bat** in una cartella (p.e. *C:\Users\<utente>\Anaconda3\Script*) contenente queste righe:

@REM Adapt the folder in the PATH to your system @SET PATH=C:\Users\Alfredo\Anaconda3;%PATH% @CALL activate py36_knime || ECHO Activating python environment failed @python %*

¹ Queste istruzioni sono si trovano in <u>https://docs.knime.com/2018-12/python_installation_guide/python_installation_guide.pdf</u>



Integrazione con Python - 2/2

5. Indicare nelle preferenze il perce dell'installazione di Python

	📩 Preferences	— 🗆 X
	type filter text	Python 🗘 👻 🕁 👻
ercorso	JavaScript Views KNIME Explorer > KNIME GUI KNIME Store Master Key Meta Info Preferen Network	See the <u>FAQ</u> for details on how to use a start script Python 2 Path to Python 2 executable python
KNIME Analytics Platform	Open Street Map Palladian Geocode Palladian Location	Could not find python executable at the given location: python. Use as default
New Cr Save C Save C	Palladian Cocation Palladian Ranking : Palladian Web Sear	Python 3 (Default)
Save All Ctrl+Sh Close All Ctrl+Shi	ft Preferred Renderer	Path to Python 3 execttable C:\Users\Alfredo\Anaconda3\py36.bat
Print Import KNIME Workflow Koport KNIME Workflow Switch Workspace	tr Report Designer > Textprocessing > Workflow Coach	Python version: 3.6.8 Use as default
Preferences Export Preferences Import Preferences	Plug-in Development Report Design Run/Debug Team	Serialization Library Flatbuffers Column Serialization
Install KNIME Extensions Update KNIME	Validation Validation	Restore <u>D</u> efaults <u>Apply</u>
Restart Exit		Apply and Close Cancel

N.B. Non tutti i package di Python sono presenti nell'integrazione. Per poterne utilizzare altri bisogna installarli. Per esempio, per il package scikit-learn bisogna prima installarlo in Anaconda e poi in Knime. Aprire il prompt di Anaconda ed eseguire questi comandi:

conda install scikit-learn oppure pip install -U scikit-learn conda install -n py3_knime scikit-learn



Gestione del Workflow

È possibile personalizzare il workflow

aggiungere annotazioni





allineare i nodi orizzontalmente/verticalmente



Welcome to KNIME Analytics Platform

D•

🛕 *0: Churn Analysis 🔀

Gestione del Workflow

cambiare l'apparenza



Riferimenti

Software

Knime	https://www.knime.org/
R	https://www.r-project.org/
python	https://www.python.org/

Bibliografia

Corsi di auto-apprendimento:https://www.knime.com/knime-self-paced-coursesLibrihttps://www.knime.org/knimepress

Forum Richieste di supporto

https://forum.knime.com/





Introduzione al software KNIME Analytics Platform Esercizi - 1ª Parte

- Avviare Knime dalla sua icona
- Creare un nuovo workflow (p.e. Radio Audience)
- Importare [Nodo File Reader] dalla chiavetta USB, o dalla directory dove sono stati scaricati i file del corso, il file *Questionario.txt* (impostare la virgola come delimitatore e i nomi delle colonne in riga 1)
- Importare [Nodo Excel Reader (XLS)] dalla chiavetta USB, o dalla directory dove sono stati scaricati i file del corso, il file Tabella Preferenze.xlsx (impostare i nomi delle colonne in riga 1)
- Unire [Nodo Joiner] in left join le 2 tabelle (*Questionario* nella porta in alto e *Preferenze* in quella in basso) usando <*risposta>* come colonna di match, che non sarà portata in uscita
- Selezionare le righe [Nodo Row Filter] della colonna <preferenza> escludendo il valore "Altro"
- Creare una Regola [Nodo Rule Engine] per la nuova colonna <fascia oraria> che avrà il valore:
 - "1. Mattino" se la colonna *<orario>* ha i valori "06-09","09-12","12-
 - 13" (ad esempio: **\$orario\$ IN ("06-09", ...) => "1. Mattino"**)
 - "2. Pomeriggio" se la colonna <orario> ha i valori "13-16","16-18"
 - "3. Sera" se la colonna <orario> ha il valore "18-20"
 - "4. Notte" se la colonna <orario> ha il valore "20-02"
- Creare una tabella di frequenze incrociate [Nodo Crosstab] dove <preferenza> è la variabile di riga e <fascia oraria> è la variabile di colonna visualizzando solo le frequenze e le percentuali di riga.

Esercizio





L'azione va incontro all'insuccesso anche perché non di rado le conoscenze radunate con fervore di zelo non erano guidate da un filo conduttore. Non conosce chi cerca, bensì colui che sa cercare.

Luigi Einaudi

Contatti <u>alfredo.roccato(at)fastwebnet.it</u> <u>www.alfredoroccato.it</u>