

[L1-DS] KNIME Analytics Platform データサイエンティスト : ベーシック

著者: KNIME AG翻訳: インフォコム株式会社

アジェンダ

1. KNIME 概要

KNIMEの概要説明、インストール / セットアップ、各画面の説明 / 基本操作方法

2. データの入力

ファイルとデータベースへのアクセス方法

3. データの操作

データの置換、結合、集計

4. ワークフローの組織化とドキュメント作成

注釈/ラベルの編集、メタノード化、

5. データの可視化

散布図/棒グラフ/折れ線グラフなど、ビューの説明

6. データマイニング

決定木/回帰分析/モデルの評価/クラスタリング(k-Means)など、ビューの説明

7.外部ツールとの結合

Java/R/Python/WEBサービスとの連携

8. データ出力とデプロイメント

各種ファイルへの出力



2

概要 KNIME Analytics Platform

KNIME Analytics Platform とは?

- データを分析、加工、可視化、レポーティングするためのツール
- グラフィカルプログラミングの枠組みに基づく
- 多様な拡張機能を提供:
 - テキストマイニング
 - ネットワークマイニング
 - ケムインフォマティクス
 - 様々なツールとの統合

 (Java、R、Python、Weka、
 Keras、Plotly、H2O など)



KNIME ワークフローを見てみよう

ノード データに関するタスクを実行











- データベース
 - MySQL, PostgreSQL, Oracle
 - Theobald
 - JDBC経由 (DB2, MS SQL Server)
 - Amazon DynamoDB

ファイル

- CSV, txt, Excel, Word, PDF
- SAS, SPSS
- XML, JSON, PMML
- 画像、テキスト、ネットワーク
- その他
 - Twitter, Google
 - Amazon S3, Azure Blob Store
 - Sharepoint, Salesforce
 - Kafka
 - REST, Webサービス



ビックデータ



- Spark & Databricks
- HDFS support
- Hive
- Impala
- In-database processing



データ変換



- 前処理
 行、列、表ベース
- データの融合
 結合、連結、付加

集約
グルーピング、ピボッティング、ビニング

• 特徴作成および選択

8

分析とデータマイニング



回帰 線形、ロジスティック

クラス分析 決定木、アンサンブル学習、SVM、多層パーセプトロン、 ナイーブベイズ

クラスタリング K平均法、DBSCAN、階層型

> 検証 クロスバリデーション、スコアリング、ROC

ディープラーニング Keras, DL4J

• ツール R、Python、Weka、H2O、Keras



可視化





対話型の可視化

- JavaScriptベースノード
 - 散布図、箱ひげ図、折れ線グラフ
 - ネットワーク図、ROC曲線、決定木
 - Plotly 連携
 - ノード続々開発中!
- Misc
 - タグクラウド、open street map(地図)、分子構造
- スクリプトベースの可視化

R, Python





- データベース

- ファイル
 - Excel, CSV, txt
 - XML
 - PMML
 - 出力先:ローカル環境、KNIMEサーバ、 Amazon S3, Azure Blob Store

BIRTレポーティング

11

ノードの数は2000以上!





KNIME Analytics Platformのインストール

- インストールするKNIMEのバージョンを選ぶ:
 - Mac, Windows 32 or 64 bit, Linux
- アーカイブファイルをダウンロードして展開、またはインストーラをダウンロードし て実行する

Windows		
KNIME Analytics Platform for Windows (installer)	32 Bit	(393.38 MB)
The installer adds an icon to the desktop and suggests suitable memory settings	64 Bit	(396.38 MB)
KNIME Analytics Platform for Windows (self-extracting archive)	32 Bit	(396.87 MB)
The self-extracting archive only creates a folder holding the KNIME installation	64 Bit	(400.72 MB)
KNIME Analytics Platform for Windows (zip archive)	32 Bit 64 Bit	(466.11 MB) (470.07 MB)

		Linux		
KNIME Analytics Platform for Linux			64 Bit	(417.21 MB)

	Мас		
KNIME Analytics Platform for Mac OSX (10.11 and above)		64 Bit	(388.44 MB)



KNIME Analytics Platform の起動

インストーラが作成したショートカット



またはインストールディレクトリでknime.exe を実行する

→ I Application Tools	KNIME_v3			- 🗆	×
File Home Share View Manage					~ 🕐
	v3		v C	Search KNIME_v3	Q
Downloads	A Name	Date modified	Туре	Size	^
🖳 Recent places	Configuration	27-Jan-16 11:58 AM	File folder		
Boxcryptor	🔐 dropins	27-Oct-15 1:36 PM	File folder		
😌 Dropbox (KNIME)	Features	27-Jan-16 11:58 AM	File folder		
	🌗 jre	27-Jan-16 11:58 AM	File folder		
🜏 Homegroup	🌗 p2	27-Jan-16 11:58 AM	File folder		
	🌗 plugins	27-Jan-16 11:58 AM	File folder		
🖳 This PC	🔼 artifacts.xml	27-Oct-15 1:39 PM	XML File	105 KB	
📙 Desktop	eclipsec.exe	16-Oct-15 7:07 AM	Application	18 KB	
Documents	🛆 knime.exe	16-Oct-15 7:17 AM	Application	312 KB	
🚺 Downloads	knime.exe.manifest	15-Oct-15 11:00 AM	MANIFEST File	1 KB	
Music	🛍 knime.ini	27-Oct-15 1:39 PM	Configuration sett	1 KB	
Dictures	knime-workspace.zip	13-Oct-15 5:01 PM	Compressed (zipp	34 KB	
J Videos	LICENSE.TXT	12-Oct-15 3:29 PM	Text Document	1 KB	
Windows8_OS (C:)	 Description Description 	12-Oct-15 3:29 PM	Adobe Acrobat D	899 KB	~
17 items 1 item selected 311 KB					:== 🔊



KNIME Workspace

- 実行中のKNIMEセッションでワークフローやデータを保存するフォルダ。
- 他システムに簡単に移植できる(KNIMEと同様)

	Analytics Platform Launcher	>
Select a dir	irectory as workspace	
KNIME Ana	alytics Platform uses the workspace directory to store its preferences and development	t artifacts.
Workspace:	C:\Users\knime\knime-workspace	Browse
Use this a	as the default and do not ask again	
Recent W	Norkspaces	
		Consel
	Launch	Cancel

KNIME Analytics Platform Workbench



16



KNIME Explorer



- ローカル環境のワークフローへのアクセス
- マウントポイントでの接続先
 - EXAMPLE Server
 - KNIME Hub
 - KNIME Server
- Explorer上部のツールバーには検索窓と以下のボタン
 - ◆ Workflow Editor で表示しているワークフロー をツリーから選択
 - ŀ ビューをリフレッシュ
- 4 種類のコンテンツを格納 ワークフロー
 ワークフローグループ
 - データファイル
 - 共有コンポーネント

ワークフローの作成、インポート、エクスポート

- KNIME Explorer内で右クリックして、新規ワークフローやワークフローグループを作成 したり、ワークフローをインポートしたりすることができます。
- エクスポートする場合、ワークフローまたはワークフローグループを右クリック



LOCAL (Local Wo BigData Trainii KNIMEUserTra	New KNIME Workflow New Workflow Group	
► data	R Import KNIME Workflow.	
A 01, impo	Export KNIME Workflow.	
02. Data 03. Visu 04. Data	X Delete A Rename	
05. Flow 06. Expt 07. Wor 08. Advi 09. Mod 10. Date	Configure Execute Cancel execution	
A 11. Inter A 12. Dep	Recution	
predictic solutions	Workflow Credentials Workflow Variables	
Final workf	Edit Meta Information	
Node Repository	🖑 Refresh	
10 No	Copy Location	•
Manipulation Views Analytics	 ✓ Cut E Copy Paste 	жx ЖС ЖV





Node Repository



 すべてのKNIMEノードがリストアップ されています。

検索ボックスには2つのモードがあります。

🗣 標準検索 - ノード名の完全一致

♀ ファジー検索 - 最も類似したノード名を検索します。

Description

Description 🛛		
	Row Filter	^
The node allows can include or e with a certain in selectable colur configure the no doesn't change t lower bounds or adapted, even if out.	s for row filtering according to certain criteria exclude: certain ranges (by row number), row row ID, and rows with a certain value in nn (attribute). Below are the steps on how ode in its configuration dialog. Note: The no the domain of the data table. I. e. the upper a r the possible values in the table spec are n one of the bounds or one value is fully filter	a. It ws 1 a 7 to ode and not red
Dialog Optio	ons	
In- or exclude ro	ows by criteria	
You must firs filtering from include or exe	t select which criteria should be used for the left-hand side. Also choose whether to clude rows according to the selected criteria	

Depending on the choice, you will then have to adjust the filter parameters in the right-hand panel.

Column value matching

If filter by attribute value is selected, select the name of the

Description ウィンドウには、以下の 情報が表示されます。

- ノードの機能
- 入力と出力
- ノードの設定
- ポート
- 文献の参照

Workflow Description



ワークフローを選択すると、
 Description ウィンドウにワークフローの情報が表示されます。

- タイトル
- 説明
- 関連タグおよびリンク
- 作成日
- 作成者

Workflow Coach

ノードのレコメンデーションエンジン

- ワークフローの中で次にどのノードを配置するかのヒントを与える
- KNIMEコミュニティの利用統計とKNIME独自の利用情報に基づいています。



Node Monitor

- デフォルトでは、Node Monitor は、ワークフローエディターで選択されたノードの出力 テーブルを表示します。
- 右上の3つのドットをクリックすると、フロー変数や設定などが表示されます。

E Console	🔥 Node M	Ionitor 🔀									8 🗖 🗖
									Show Output	Table	
Node: Ge	et Customers	from Data	base (0:1	207)					Show Variable	es	
State: EX	KECUTED								Show Configu	iration	
									Show Entire C	configuration	
Port Outp	ut F	Port 0		Load data					Show Node T	iming Informat	ion
									Show Graph A	Annotations	
ID		MaritalSt	atus Gend	er Estimated Yearly	Income NumberOf	Contracts Age	Available	401K CustomerV	onon orapin,		oducts
CustomerII	D: 722204	S	F	80000	4	42	1	1	4	5	Private Investn
Customeril	D: 489847	М	М	60000	2	46	1	1	4	3	Private Investn
CustomerII	D: 8444723	М	M	40000	1	32	1	2	3	0	P+B Investmer
CustomerII	D: 1487427	М	М	30000	2	63	1	1	2	2	P+B Investmer
CustomerII	D: 4693433	М	M	20000	2	63	1	1	3	4	Gold Investme
CustomerII	D: 7724940	М	М	30000	2	33	1	2	3	0	P+B Investmer
CustomerII	D: 9784443	M	M	60000	2	34	1	2	3	0	P+B Investmer
CustomerII	D: 3177757	м	М	70000	2	57	1	1	5	2	Fund Manager



Console とその他の View

E Console 🔀		🗎 🔝 🖻 📑 🚍 - 📑 🗖
KNIME Console		
*****	******	******
*** Welcome to the KNIME Anal	ytics Platform v3.3.2.v201704041558	***
*** Copyright by KNI	ME GmbH, Konstanz, Germany	***
*****	******	******
Log file is located at: /Users/kath	rinmelcher/knime-workspace/.metadata/	/knime/knime.log
WARN File Reader 2:232	Can't access 'knime://knime.workfl	ow///data/SAS_Scores.csv'. (/User
WARN File Reader 2:236	Can't access 'knime://knime.workfl	ow///data/Webdata%200ld%20System.
WARN PMML Reader 2:244	Input file '/Users/kathrinmelcher/	/knime-workspace/New_ETL_Course/data/U
WARN Database Reader 2:238	Could not determine table spec fro	om database query: [SQLITE_ERROR] SQL

Show View

View Node Help

コンソールは、内部で何が起こっているかについてのエラーや警告メッセージを表示します。

[View]をクリック、[Other...]を選択 すると、さまざまなビューが追加され ます。

Node Monitor、ライセンス、など

ノードの挿入と接続

- ノードをWorkspace上で挿入するには、Node Repository からノードをドラッグするか、 Node Repository 内でノードをダブルクリックします。
- ノードAの出力ポートを左クリックし、カーソルをノードBの(一致する)入力ポートに ドラッグしてノードを接続する





ノードの詳細

ノードには4つの状態が存在:

File Reader



未設定:

ノードは設定または受信データを待っている。

File Reader



設定済: ノードが正しく設定され、実行できる。

File Reader

実行済: ノードが正常に実行された。 結果を下流のノードで表示、使用できる。

File Reader



エラー: ノードの実行中にエラーが発生。



ノード設定

■ ほとんどのノードで設定が必要	•••		Dialog - 0:2 - File Re	ader
	Setti	gs Flow Varia	bles Job Manager	Selection
	Enter ASCII data	file location: (pre	ss 'Enter' to update	preview)
	knime://knime	workflow///d	ata/iris.csv	
		Preserve user	settings for new lo	ration R
ノードの設定画面を表示するには	- Pasic Settings		Sectings for her lot	
	Basic Settings	C .1	1.1.1	
■ ノードをダブルクリック	read row IDs	Col	umn delimiter: ,	Y
	🔽 read column	headers 🛛 🗸	ignore spaces and t	abs
■ 右クリック -> 設定			lava-style commen	ts Sinal
	Preview			
	Click colum	n header to chan	ge column propertie	es (* = name
	Row ID D S	epal D Sepal	D Petal D Pe	tal S Sp
	1 5.1	3.5	1.4 0.2	setosa
	2 4.9	3	1.4 0.2	setosa
	3 4.7	3.2	1.3 0.2	setosa
	4 4.6	3.1	1.5 0.2	setosa
	5 5	3.6	1.4 0.2	setosa
	0 5.4	3.9	1.7 0.4	setosa
	7 4.0 9 5	3.4	1.4 0.3	setosa
	9 44	2.9	1.5 0.2	setosa
	10 49	3.1	1.4 0.2	setosa
	11 5.4	3.7	1.5 0.2	setosa
	12 4.8	3.4	1.6 0.2	setosa
	13 4.8	3	1.4 0.1	setosa
	14 4.3	3	1.1 0.1	setosa
	15 5.8	4	1.2 0.2	setosa
	16 5.7	4.4	1.5 0.4	setosa
	17 5.4	3.9	1.3 0.4	setosa
	18 5.1	3.5	1.4 0.3	setosa
			ОК	Apply



Memory Policy

Browse...

• •=?



?

Cancel

ノード実行

ノードの実行方法

- ノードを右クリック。
- メニューから [Execute] を選択。
- 実行が成功した場合、ステータスは緑色に点灯。
- エラーが発生した場合、ステータスが赤で点灯。

File Re	ader			
	Ð	Configure	F6	
	0	Execute	F7	
		Execute and Open Views	Shift+F10	
	0	Cancel	F9	
		Reset	F8	
	=	Edit Node Description	Alt+F2	
	≡Ĵ	New Workflow Annotation		
	2	Connect selected nodes	Ctrl+L	
	2	Disconnect selected nodes	Ctrl+Shift+L	
	÷	Create Metanode		
	4	Create Component		
	⊒?	Compare Nodes		
		Show Flow Variable Ports		
	ot	Cut		
		Сору		
	Ē	Paste		
	\checkmark	Undo		
	5	Redo		
	×	Delete		
	₽	File Table		



📑 🖌 🔚 🖳 😳 🖄 🚺 100% 🔽 🕴 🚥 🖧 🔂 🖸 😳 😳 💿 🐼 🖾 💻 🗮 🔍 🕼 🕼 🕼 🔘 💞 🌾 🖒

ツールバーのボタンは、アクティブなワークフローで使用することができます。

最も重要なボタンは

選択された実行可能なノードを実行(F7)

実行可能なすべてのノードを実行

選択されたノードを実行し、ファースト・ビューを開く。

選択された実行中のノードをすべてキャンセル (F9) PR 1

実行中のノードをすべてキャンセル





- ノードを右クリックすると、以下の方法で実行結果を確認 できます。
 - 出力ポート(コンテキストメニューの最後のオプション)を選択して、表や 画像などを確認する。
 - インタラクティブ・ビューを選択すると、可視化された結果がブラウザーで 表示されます。



Ð	Configure	F6							
0	Execute	F7							
8	Execute and Open Views	Shift+F10							
Θ	Cancel	F9							
Ξ	Reset	F8							
F	Edit Node Description	Alt+F2							
≡Ĵ	New Workflow Annotation								
2	Connect selected nodes	Ctrl+L							
R	Disconnect selected nodes	Ctrl+Shift+L							
÷	Create Metanode								
<u>\</u>	Create Component								
Q	Interactive View: Scatter Plot								
⊒?	Compare Nodes								
	Show Flow Variable Ports								
of	Cut								
	Сору								
Ē	Paste								
\checkmark	Redo								
5									
×	Delete	Data View							
Ð	Image								
EQ.	Input data and view selecti	on							



KNIME ファイルの拡張子

KNIME Analytics Platformに関連する

ワークフローおよびワークフローグループのための専用ファイル拡張子





はじめに: KNIME Hub





はじめに: KNIME Example Server

KNIME Explorer経由でアクセスできる公開リポジトリ

- 様々なワークフロー例がダウンロードできる
- KNIME Hub経由でも利用可能





Ð $\sim \parallel$

Open for Innovatio

KNIME

A KNIME Explorer 🐹

ショートカットキー (参考)

Task	Hot key	Description		
ノード設定	F6	選択したノードの設定画面を開く		
	F7	選択された設定ノードを実行		
	Shift + F7	設定されたすべてのノードを実行する		
ノード実行	Shift + F10	設定されたすべてのノードを実行し、すべてのビューを開く		
	F9	選択された実行中のノードをキャンセルする		
	Shift + F9	実行中のすべてのノードをキャンセルする		
/ ド 控 结	Ctrl + L	選択したノードを接続		
	Ctrl + Shift + L	選択したノードを切断する		
/ド22動レ	Ctrl + Shift + Arrow	選択されたノードを矢印の方向に移動させます。		
アノテーション	Ctrl + Shift + PgUp/PgDown	選択したアノテーションを、重なっているすべてのアノテーションの 前または後ろに移動させる		
	F8	選択したノードをリセットする		
ロークフロー堝佐	Ctrl + S	ワークフローの保存		
	Ctrl + Shift + S	開いているワークフローをすべて保存		
	Ctrl + Shift + W	開いているワークフローをすべて閉じる		
メタノード	Shift + F12	メタノード・ウィザードを開く		





今日の例題:解約予測(Churn Prediction)





今日の例題:解約予測(Churn Prediction)





- 演習で使用するデータファイルは、「data」フォルダに用意されています。さまざまなファイル形式のデータファイル、Web ベースのデータ、データベース上のデータなどがあります。
- 解約予測に使用するための顧客データは提供元の異なるデータ を混ぜたものです。
- 「Data Explorer」ノードは、データを検査するのに便利です。

	Numeric Nominal Data F	Numeric Nominal Data Preview Search:					
Explorer	Column It	Exclude Column	Minimum 🕼	Maximum 🕼	Mean \downarrow †	Standard Deviation	
	CustomerKey		11000	27336	19281.750	5319.909	
Lu 🕨 🔰	WebActivity		0	5	1.000	1.524	
	SentimentRating		0	5	1.846	1.619	
	EstimatedYearlyIncome		10000	170000	57066.921	32242.624	
	NumberOfContracts		0	4	1.493	1.145	
	Age		29	100	48.288	11.382	
	(Target		0	1	0.489	0.500	





ご不明な点やご相談などございましたら、お気軽にお問合せください。

- ・定期開催トレーニングの開催予定一覧
 <u>https://knime-infocom.jp/news/type/training/</u>
- ・トレーニングのお申し込み

https://knime-infocom.jp/trainingform/

・お問合せ先 インフォコム株式会社 ヘルスケアサービス部 KNIME担当

MAIL

Web

- TEL : 03-6866-3860
 - : <u>knime@infocom.co.jp</u>
 - : <u>https://knime-infocom.jp/contact_mail/</u>