

現場を巻き込む ものづくりデータサイエンスの進め方



株式会社KSKアナリティクス

目次

1. KSKアナリティクス会社紹介
2. データ分析人材育成の意義と目的
3. 価値を出すための分析組織
4. 機械学習ツール 「RapidMiner」
5. 機械学習トレーニング

1. KSKアナリティクス会社紹介

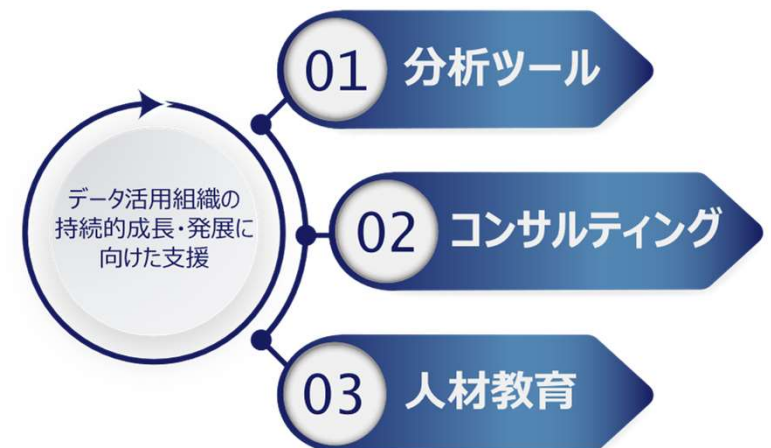
事業内容

KSKアナリティクスは、分析活用のプロフェッショナルとして
「社内に分析チームを作りたい」「データ活用を企業文化にしたい」
とお考えの企業に2つの切り口で支援をしています。

データの統合・蓄積から分析まで



3つの支援を提供



1. KSKアナリティクス会社紹介

ビジョン

～誰もが当たり前にデータを分析・活用できる社会～



一部の機械学習の専門家（分析コンサル）にデータ分析を任せていると
結果的に高コストになり、社内に分析ノウハウ（解決の道筋）が蓄積されません。
自分達自身で分析力を高めていく必要があると考えております。

1. KSKアナリティクス会社紹介

水を与えるな！井戸の掘り方を教えよ！



途上国支援PJマネージャー：

「私たちは、この井戸が完成したら次の国へ行ってしまう。ここで長く井戸を使い続けてもらうには、現地の人に修理方法まで教え、彼らが井戸を守っていかなければなりません。壊れたらおしまい、では意味がない。だから現地で入手できる部品を使い、彼らと共に井戸を作るのです。」

1. KSKアナリティクス会社紹介

機械学習への取り組み 従来モデル vs 新モデル

【従来モデル】

モデル作成はAIベンダー

多大なコミュニケーションコスト

高コスト運用

社外にノウハウが流出

【新モデル】

モデル作成は業務知識がある自分自身で
(納得できるものを作成)

自分達で素早く効率的な失敗

低コスト運用

社内に分析ノウハウが蓄積

しかし新モデルに適応するには人と時間が必要

2. データ分析人材育成の意義と目的

分析人材育成の意義

製造業の分野は幅広い

- ・製造業は数万種に至る分野があります。総合大学の工学部には十の学科があり、研究室は百を超え、研究テーマは千を超える
- ・それらの分野別に要求される専門知識は異なるので、「**分野別のデータ分析者**」が(いずれ)必要になる
- ・製造業系分析者教育の潜在的な必要性は大きく、「**製造業系分析者を育成できる人**」の需要も大きくなるものと予想される



複数人が、複数課題がこなすことができれば、少しの改善、少しの効率化が1年間通して実施される。「塵も積もれば」、多くの利益を創出することができる。社内に分析ノウハウが蓄積されていけば製品開発も加速し、結果として競争力が高まる。

2. データ分析人材育成の意義と目的

目的と到達点

製造業の分析課題（一例）

- センサーデータを用いた良品・不良品の判別（異常検知）
- 製造条件の最適化（製造工程のボトルネックの発見）
- 音響データに基づいて製品の品質予測
- 数値化された画像データを用いた製品分類（類型化）
- 品質バラツキ要因の特定
- 部品交換タイミングの最適化（予測）
- 設備ダウンタイムの予測、設備の稼働率予測
- 各種センサーの計測間隔の最適化
- 工場エネルギー需要の予測
- 製品需要予測に基づく、在庫最適化
- 各種混合データを用いた物性の予測・最適化（MI）
- 材料の形成予測 など。



人材作り：分析のための分析教育ではなく、利益を創出するための教育をすること

組織作り：年間20～30のユースケースと投資金額に見合う効果を得る仕組みを作ること

3. 価値を出すための分析組織

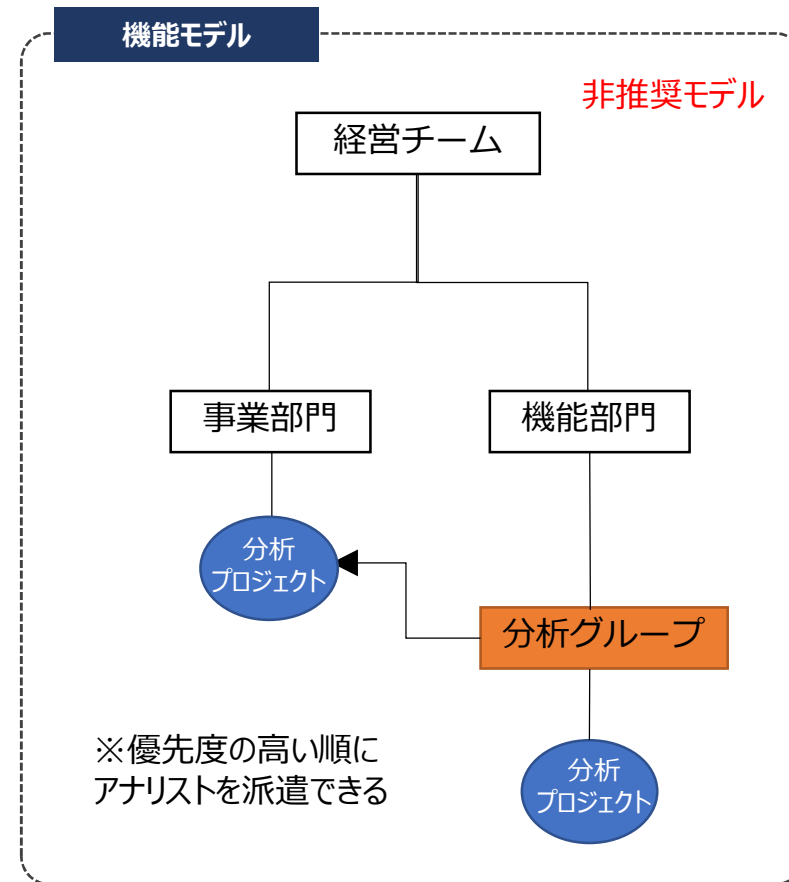
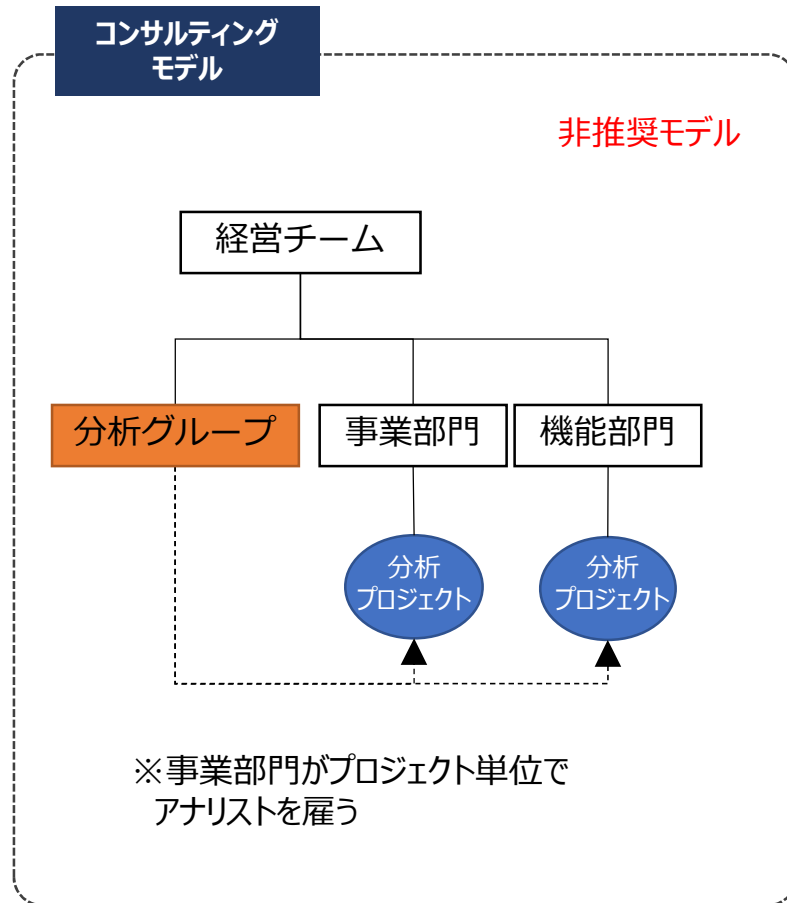
背景

- 現場と噛み合っていないデータサイエンスチームが多い
- 自動機械学習の出現により、従来の組織構造では機能しにくくなってきた。

多くの企業で組織構造を見直す必要がでてきている。

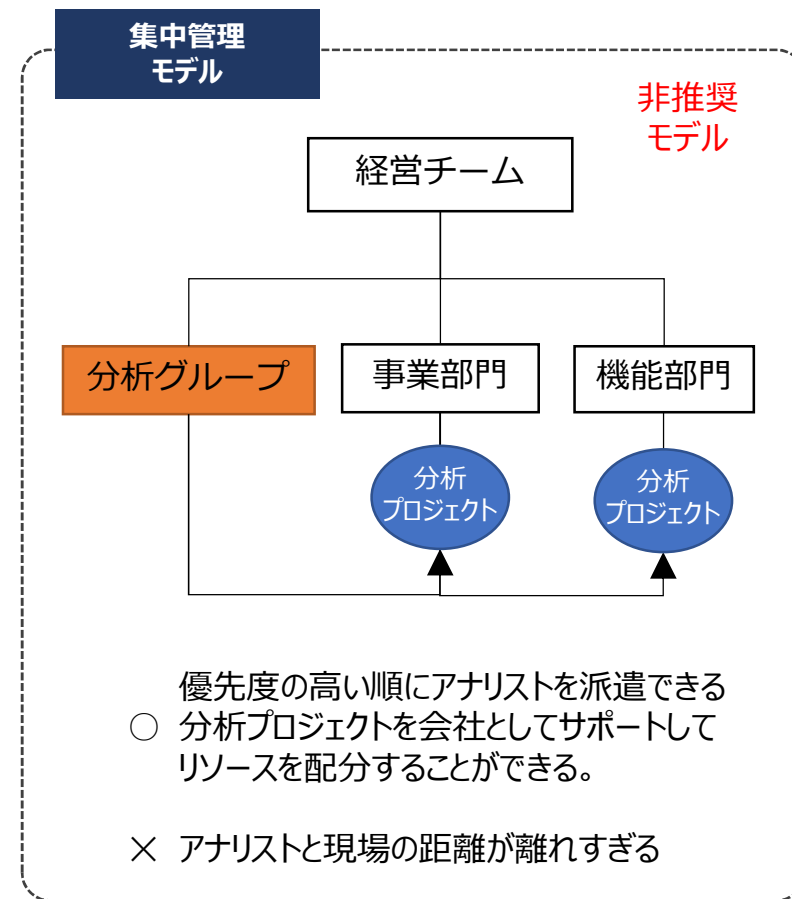
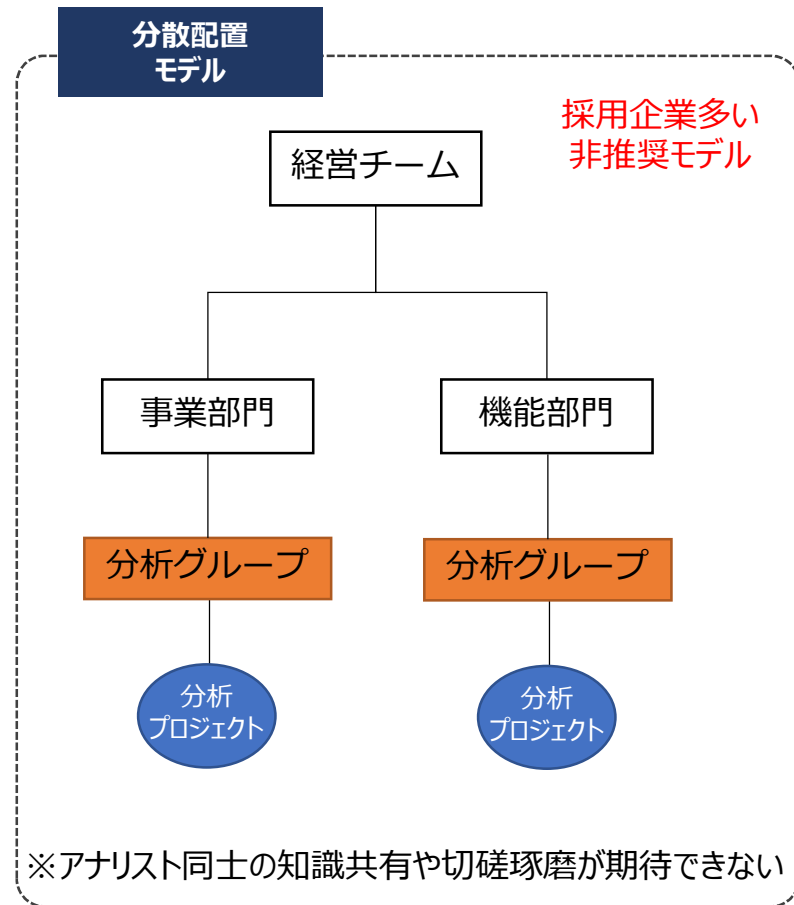
3. 価値を出すための分析組織

組織体制



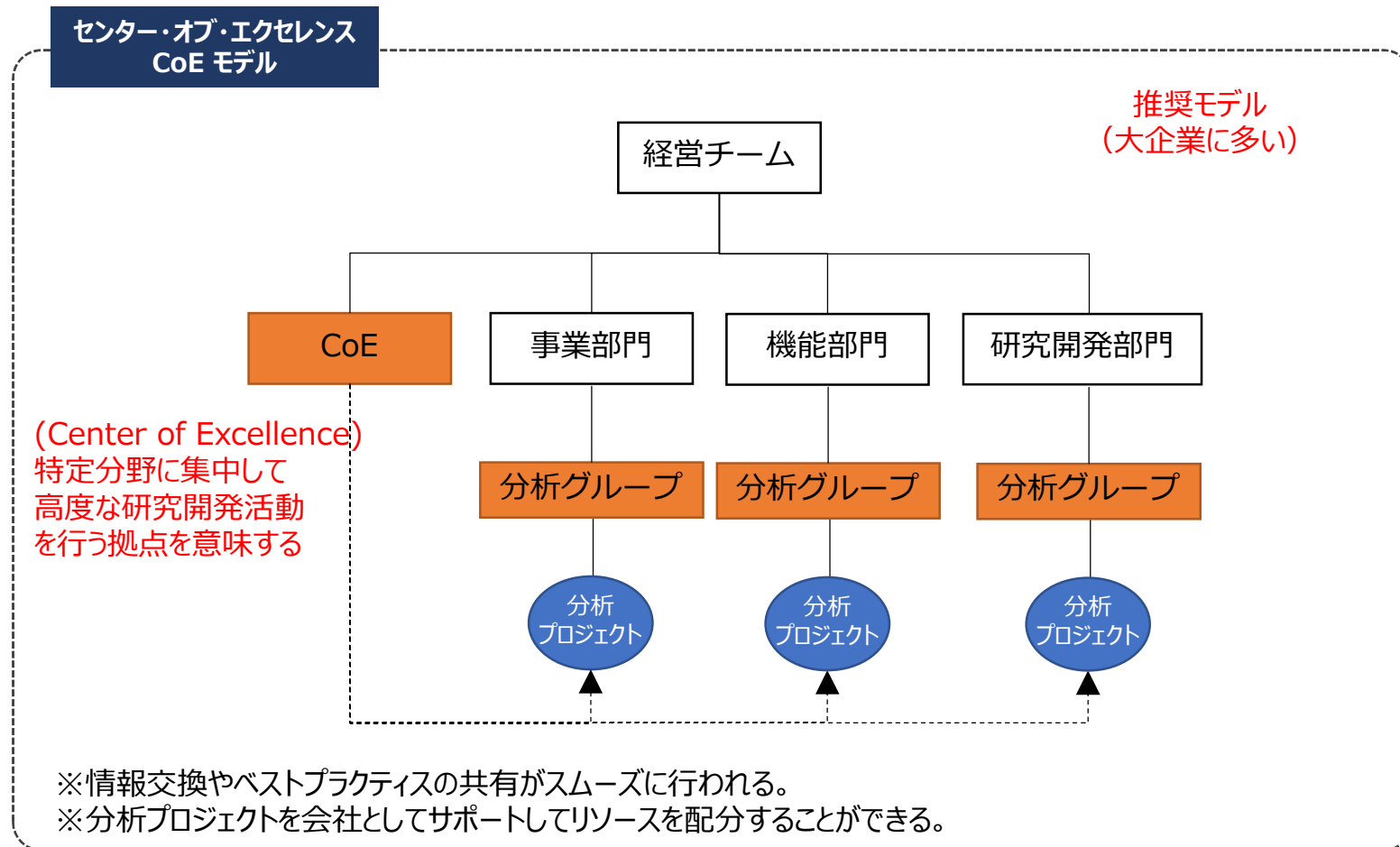
3. 価値を出すための分析組織

組織体制



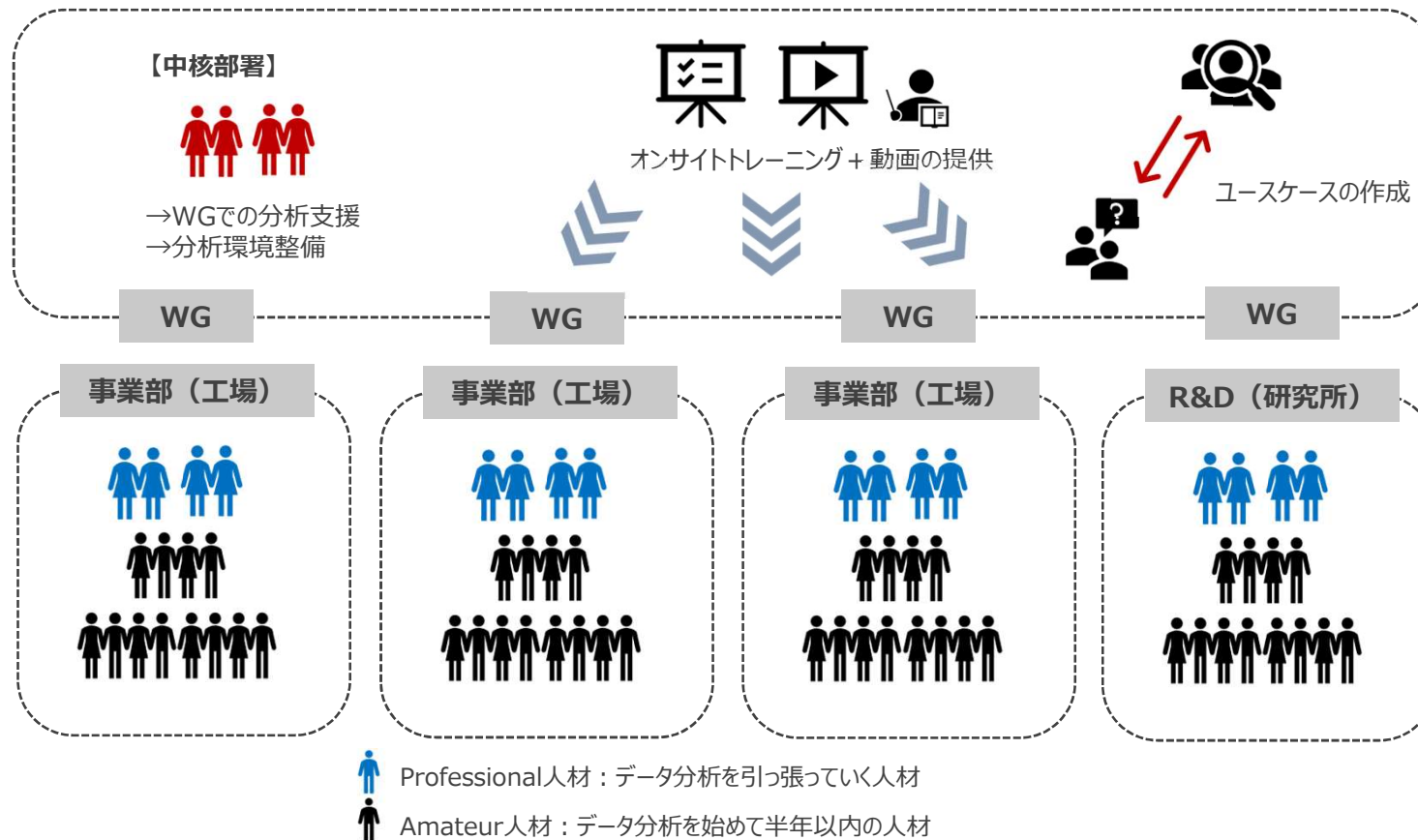
3. 価値を出すための分析組織

組織体制



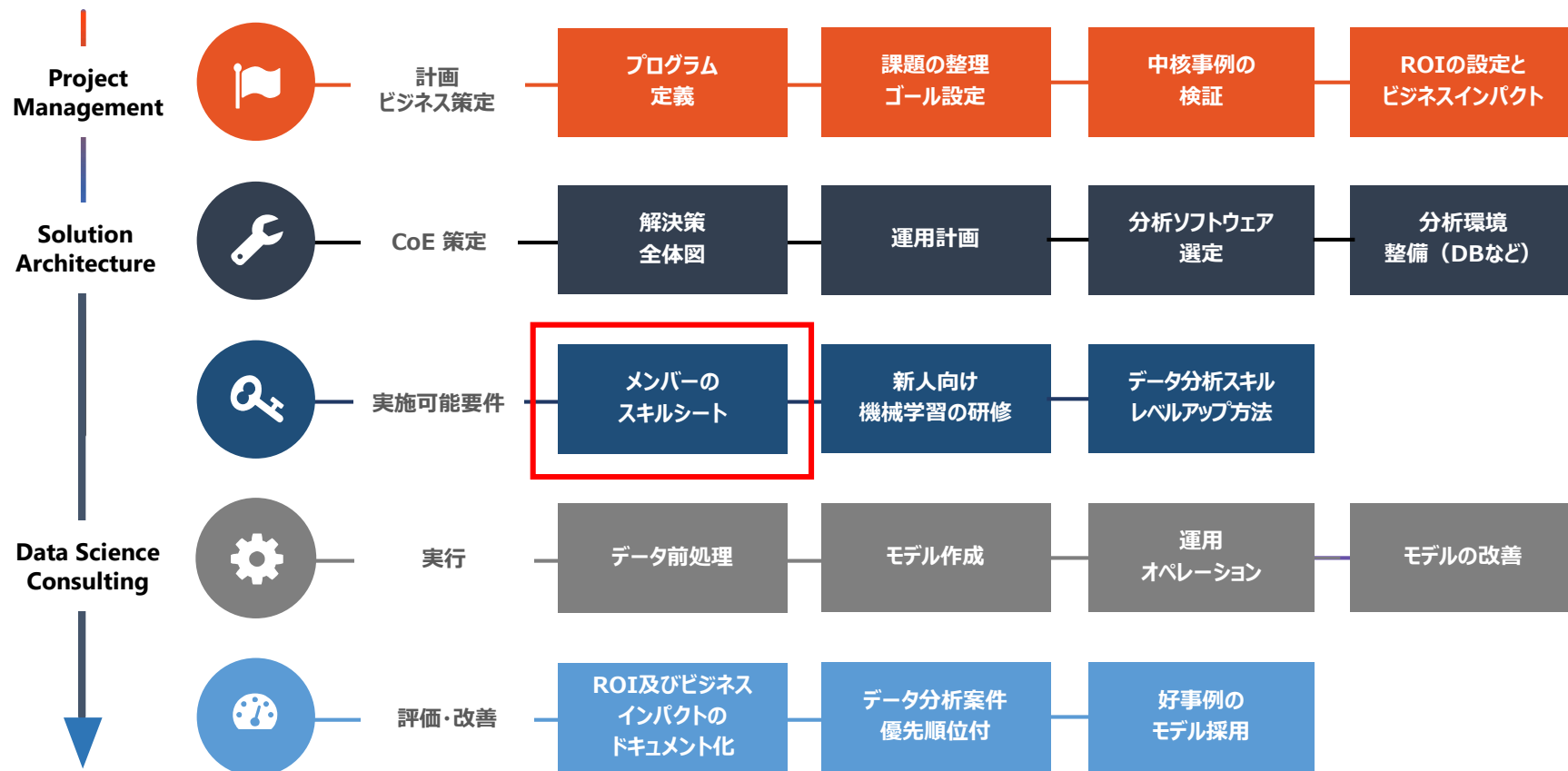
3. 価値を出すための分析組織

組織体制（某お客様のCoE組織例）



3.価値を出すための分析組織

中核部署の役割（全体図）



3. 価値を出すための分析組織

中核部署の役割（人材要件）

タレント	任務	スキル	主導	補佐
プロジェクト管理	● チームの立ち上げ、予定、スケジュール管理。 ● リソースの整理 ● トラブル解決	● 組織 ● 方法論 ● 人材管理	● プロジェクトの立ち上げと実行	● オペレーションの継続
データ収集前処理	● データの構造化 ● データ前処理のプロセスを再現できる、共有する	● システム構築 ● コーディングスキル ● 統計処理	● 分析プロジェクト発足当初。初期段階で必須	● 仮説の検証 ● データの視覚的探求
データ分析（機械学習）	● 予測モデル作成 ● パターン（知識）発見	● 統計 ● 機械学習 ● クリティカルシンキング	● データの視覚的探求 ● 非技術者とのコミュニケーション	● プロジェクトの発展段階。非専門家へのプレゼンテーション
専門知識プレゼンテーション	● ビジネス目標の定義 ● 仮説の検証と提案 ● 分析結果の解釈	● クリティカルシンキング ● 運用面の戦略	● データの可視化とプレゼンテーション ● プロジェクトのビジネスインパクト	● 可視化のデザインとビジュアル面のプロトタイプ作成

3.価値を出すための分析組織

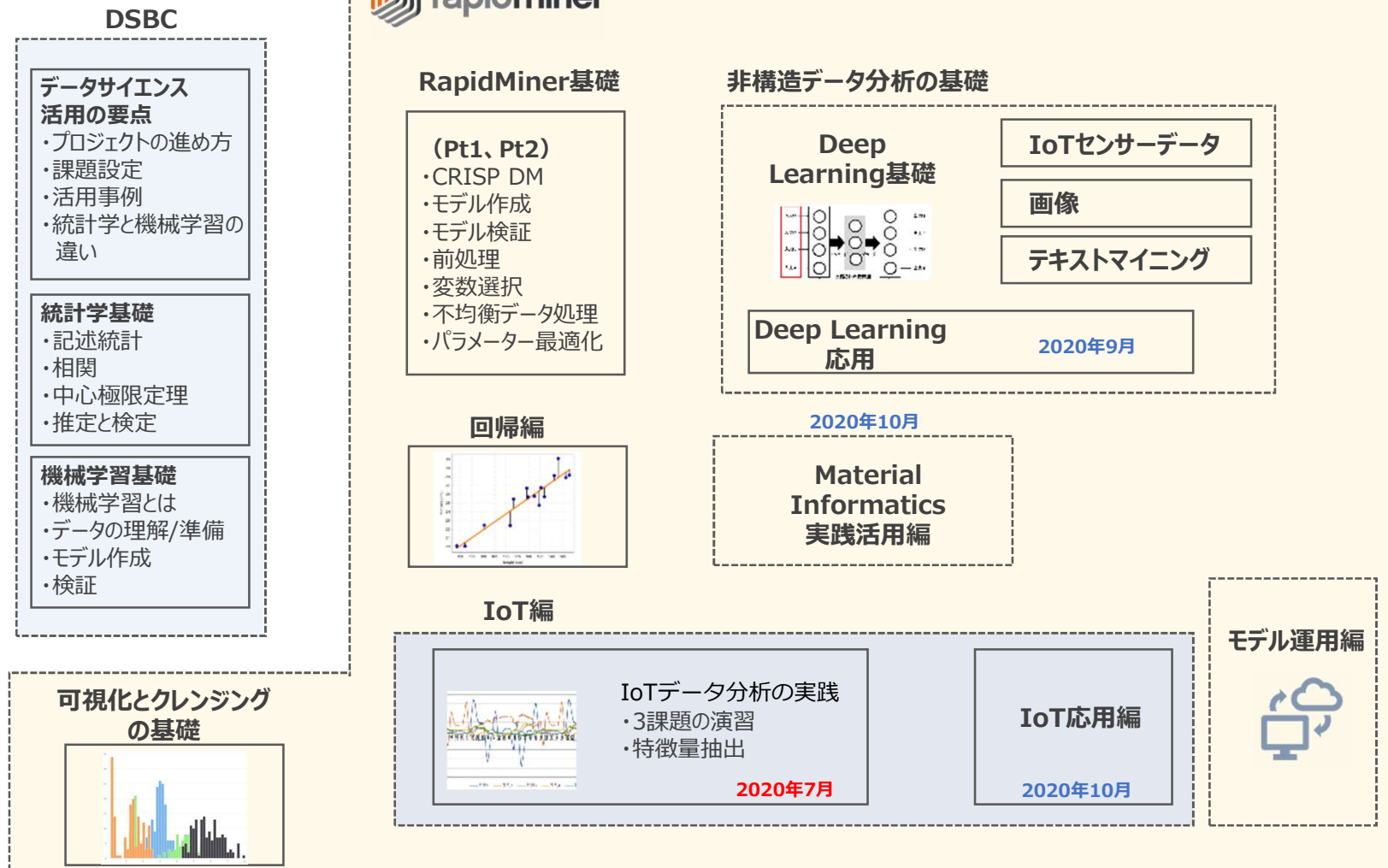
中核部署の役割（人材スキルシート）

	メンバー		プロジェクト 管理	データ収集 前処理	データ分析 (機械学習)	専門知識 プレゼンテーション	トータル スキル
Team A	◎リーダーA		5	2	4	4	15
	A-1		3	5	5	3	16
	A-2		2	5	2	2	11
	A-3		2	2	4	3	11
	A-4		2	2	4	2	10
Team B	◎リーダーB		4	3	5	4	16
	B-1		5	5	5	3	18
	B-2		4	5	4	4	17
	B-3		2	2	1	2	7
	B-4		1	1	1	2	5

※各スキルを5段階で評価

3.価値を出すための分析組織

コースマップ



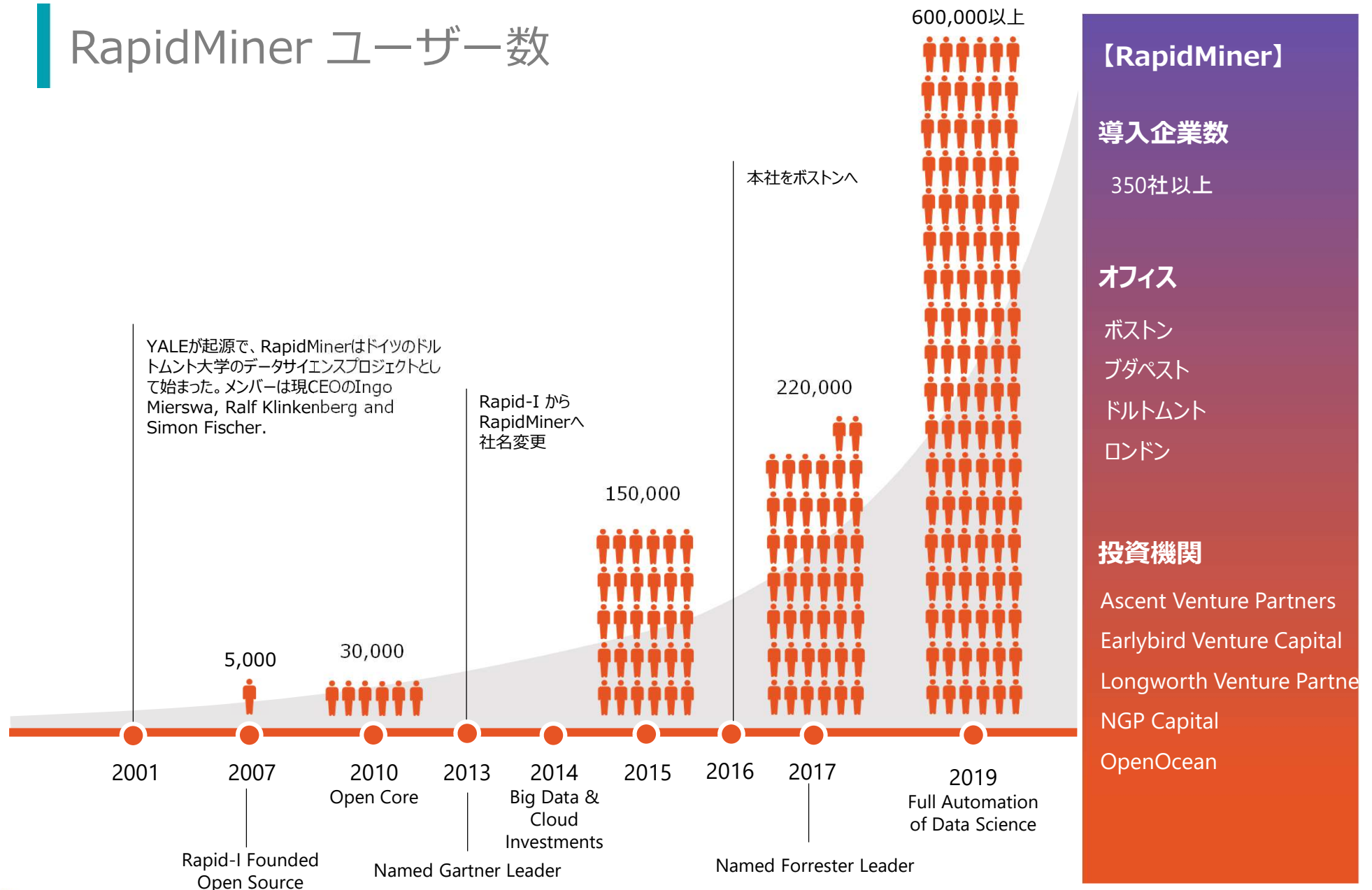
基礎

中級

応用

4. 機械学習ツール「RapidMiner」

RapidMiner ユーザー数

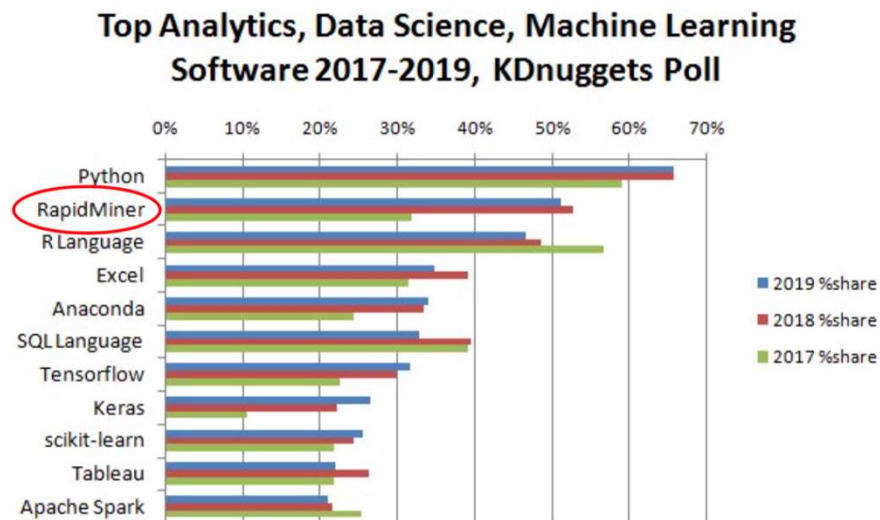


4. 機械学習ツール 「RapidMiner」

RapidMiner の評価

ワールドワイドで支持されている

世界的に有名なデータマイニング情報サイト「KDnuggets」による、過去1年間で利用したデータ分析ソフトウェア調査(2017-2019)。GUIでは最もよく利用されており、リーダーポジションにあります。



Software	2019 % share	2018 % share	2017 % share
Python	65.8%	65.6%	59.0%
RapidMiner	51.2%	52.7%	31.9%
R Language	46.6%	48.5%	56.6%
Excel	34.8%	39.1%	31.5%
Anaconda	33.9%	33.4%	24.3%
SQL Language	32.8%	39.6%	39.2%
Tensorflow	31.7%	29.9%	22.7%
Keras	26.6%	22.2%	10.7%
scikit-learn	25.5%	24.4%	21.9%
Tableau	22.1%	26.4%	21.8%
Apache Spark	21.0%	21.5%	25.5%

4. 機械学習ツール 「RapidMiner」

Python、Rを使った機械学習

```
1 from sklearn.svm import SVC
2 from sklearn.model_selection import train_test_split
3 import numpy as np
4 import pandas as pd
5
6 # class: 0
7 df_a = pd.DataFrame({'x1': np.random.randn(100),
8                     'x2': np.random.randn(100),
9                     'y' : 0})
10
11 # class: 1
12 df_b = pd.DataFrame({'x1': np.random.randn(100) + 5,
13                     'x2': np.random.randn(100) + 3,
14                     'y' : 1})
15
16 # トレーニングデータとテストデータに分割
17 X_train, X_test, y_train, y_test = \
18     train_test_split(df[['x1', 'x2']], df['y'], test_size=0.2)
```

Python

```
7 ## Creating a simple tuned model ----
8 # automated parameter tuning of C5.0 decision tree
9 set.seed(300)
10 m <- train(default ~ ., data = credit, method = "C5.0")
11
12 # summary of tuning results
13 m
14
15 # apply the best C5.0 candidate model to make predictions
16 p <- predict(m, credit)
17 table(p, credit$default)
18
19 # obtain predicted classes
20 head(predict(m, credit, type = "raw"))
21
22 # obtain predicted probabilities
23 head(predict(m, credit, type = "prob"))
24
25 ## Customizing the tuning process ----
26 # use trainControl() to alter resampling strategy
27 ctrl <- trainControl(method = "cv", number = 10,
28                     selectionFunction = "oneSE")
```

R

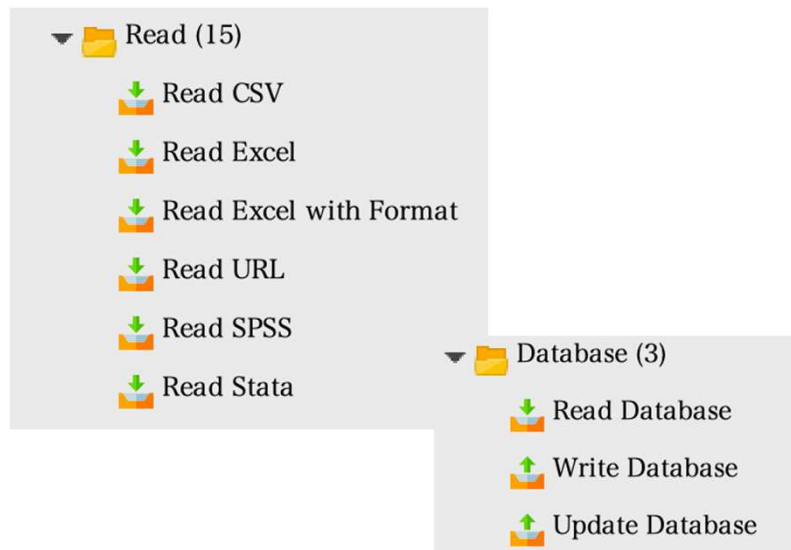
Python、Rを習得しようとする習得コスト（時間）が高く、
本来の目的に集中できない

4. 機械学習ツール 「RapidMiner」

RapidMiner Studio 総合分析プラットフォーム

① データアクセス

- 様々なファイル形式に対応
- あらゆる主要データベースにアクセス
- JDBC接続



② 可視化・統計量の表示

- データの分布、特徴を把握
- 豊富な可視化ツール（33種類）

3-D散布図、ヒストグラム、箱ひげ図
密度分布、円グラフ、バブルプロット etc.



4. 機械学習ツール 「RapidMiner」

RapidMiner Studio 総合分析プラットフォーム

③ データ加工、クレンジング

- 様々なデータ加工機能

データ統合、フィルタリング、ソート、アグリゲーション、etc.

- 特徴量の選択、特徴量の生成

自動最適化、次元圧縮、主成分分析 etc.



④ モデリング

- 150を超える分析アルゴリズムが利用可能

回帰分析、クラスタリング、決定木、サポートベクターマシン、アンサンブル学習、ディープラーニング etc.

- RapidMiner上でPythonやRのコードも実行可能。
- 拡張機能も充実している。



4. 機械学習ツール 「RapidMiner」

RapidMiner Studio 総合分析プラットフォーム

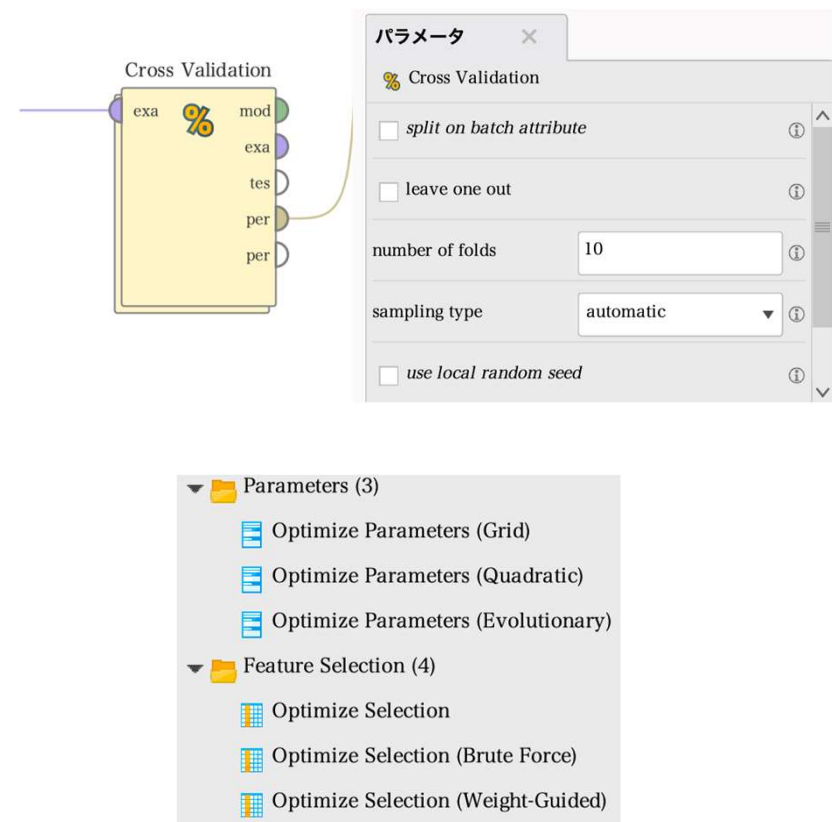
⑤ 検証、最適化

- クロスバリデーションで
モデルの精度を確認
- 様々なパフォーマンス指標

正答率(Accuracy)、適合(Precision)
再現率(Recall)、RMSE、AUC、
有意差検定、etc.

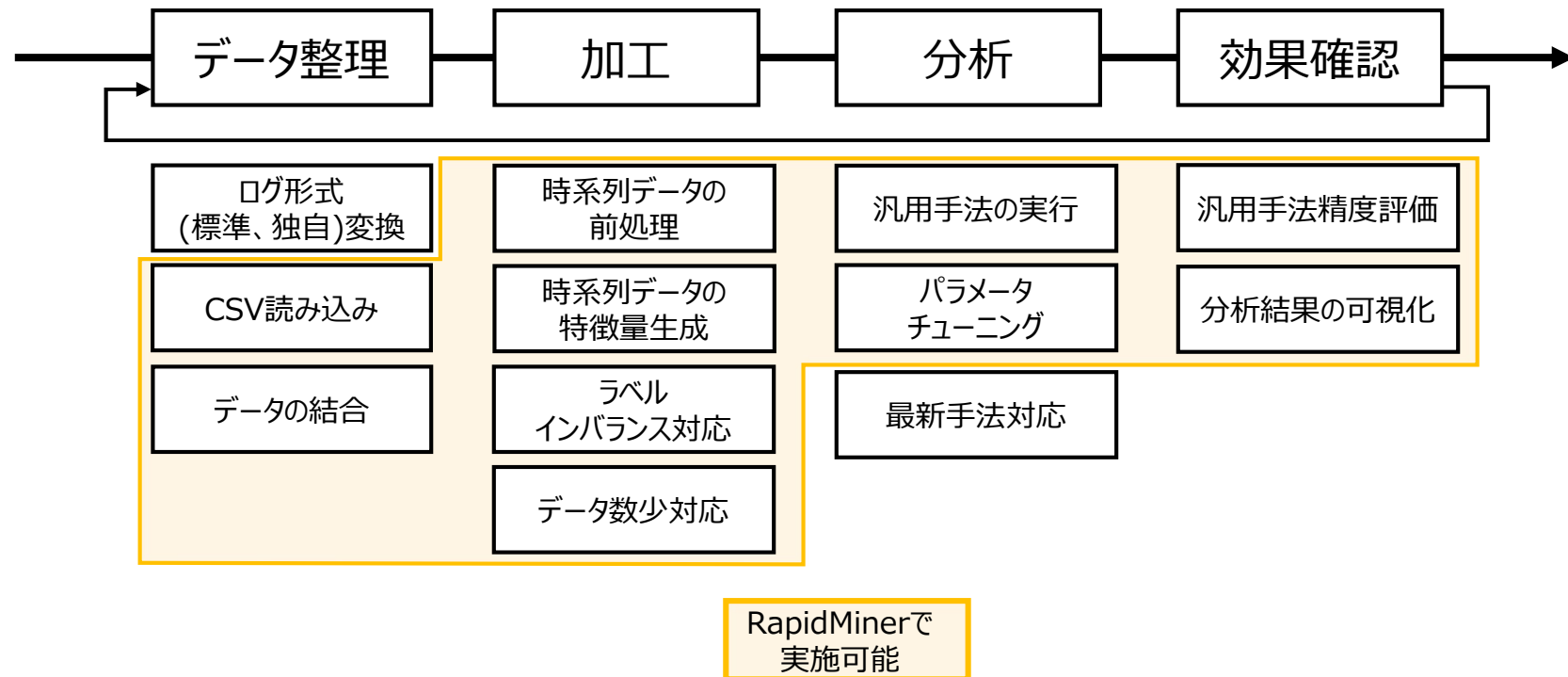
- パラメータの最適化

Grid (グリッドサーチ)
Evolutionary、etc.



4. 機械学習ツール 「RapidMiner」

RapidMiner Studio 対応領域



RapidMinerはデータ分析サイクルを早く（Rapid）回せるようになることを意識した機能（オペレーターなど）の開発にリソースを割いています。

4. 機械学習ツール「RapidMiner」

Autoモデル機能の活用

Auto Model

自動モデル生成機能

Auto Model

Load DataSelect TaskPrepare TargetSelect InputsModel TypesResults

RESTARTBACKOPEN PROCESSEXPORT

Results

ComparisonOverviewROC Comparison

Naive BayesGeneralized Linear ModelLogistic RegressionFast Large MarginDeep LearningDecision TreeRandom ForestGradient Boosted Trees

SAVE RESULTS

Overview

Total number of created models: 174

Accuracy

Runtimes (ms)

Accuracy

Model	Accuracy ↓	Standard D...	Total Time	Training Time (1...	Scoring Time (1...
Gradient Boosted Trees	81.5%	± 6.7%	20 s	1 s	-0 ms
Deep Learning	78.3%	± 4.6%	1 s	3 s	-0 ms
Support Vector Machine	76.2%	± 4.1%	2 s	40 ms	120 ms
Random Forest	75.0%	± 5.9%	6 s	928 ms	361 ms
Fast Large Margin	71.2%	± 4.3%	900 ms	8 ms	-0 ms

4. 機械学習ツール「RapidMiner」

Autoモデル機能の活用

生成したモデルでシミュレーションが可能です。
シミュレーションだけでなく、制約条件を指定し、ある確率を最も高く（低く）
するような条件を算出する最適化機能がついています。逆解析することで、
最も良い製造条件や材料の組み合わせを知る事ができます。



Data Analysis *for* Everyone!



KSK ANALYTICS