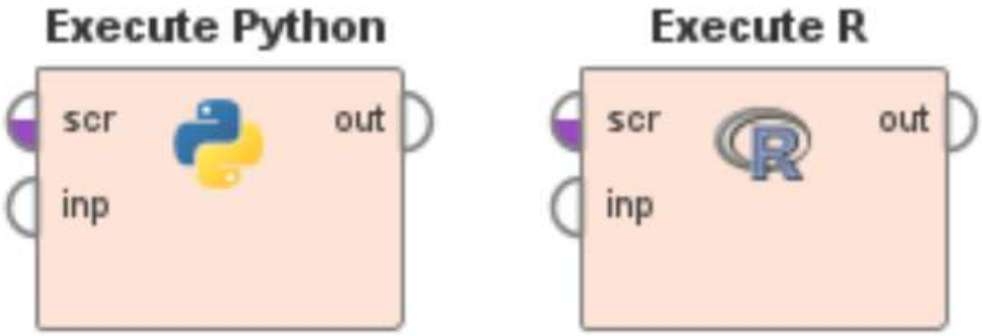
PythonとRapidMinerの
JupyterNotebook統合

株式会社KSKアナリティクス

1 .従来使われてきたPython/Rとの統合

はじめに

- RapidMinerはRとPythonとの統合に力を入れており、RapidMiner Studioでネイティブオペレーターと Execute RかPythonのオペレーターを組み合わせることで統合できます。



1.従来使われてきたPython/Rとの統合

RapidMinerとの統合メリット

- コーディングのみの機械学習にも良さがありますが、組み合わせることで新たな力を発揮します。

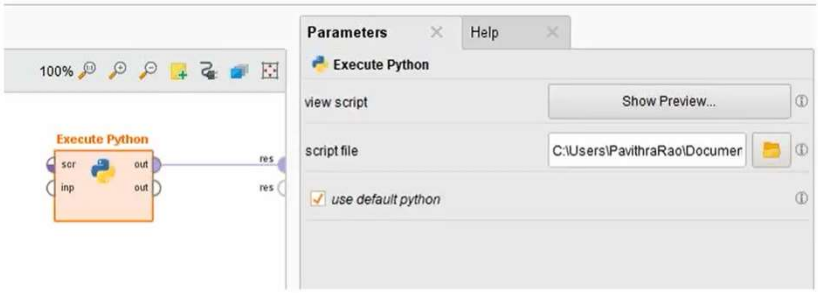
コーディングのみの良さ
・最先端のオープンソースである
・フリーである
・専門家レベルの柔軟性がある

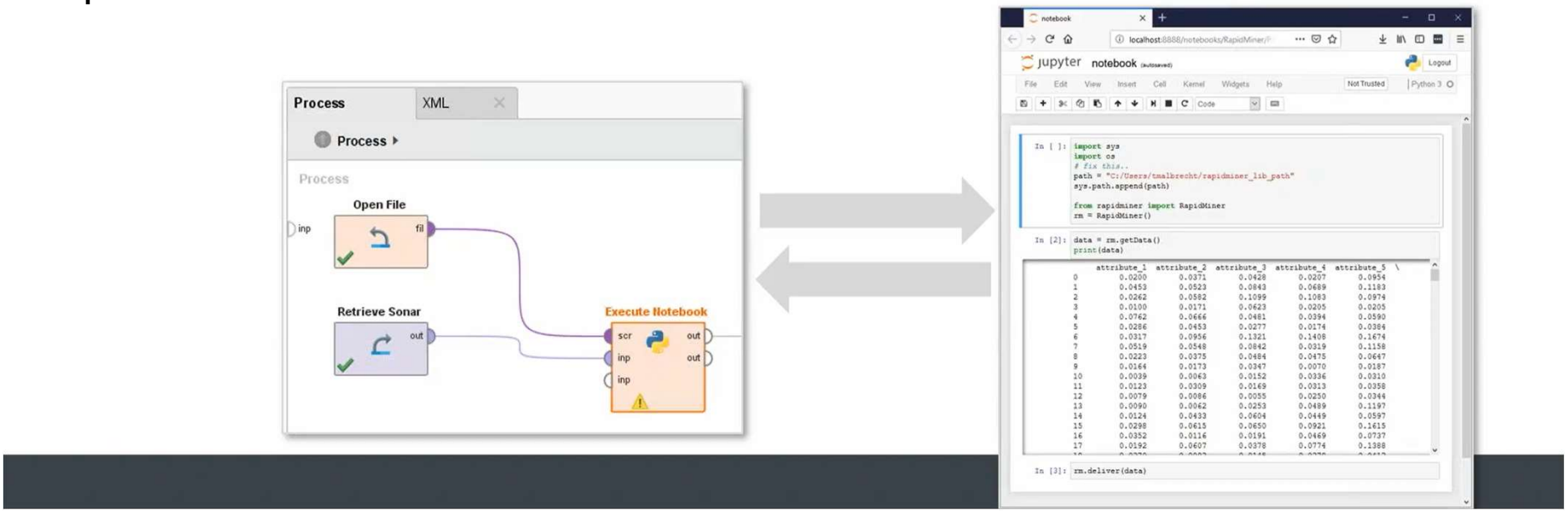
組み合わせの良さ
・試作が早い
・手作業の自動化、低スキル化
・思考の人と行動の人のコラボ
・エンタープライズレベルへのスケール化
・標準化できる

2.どのように使用するか

Studioでの利用

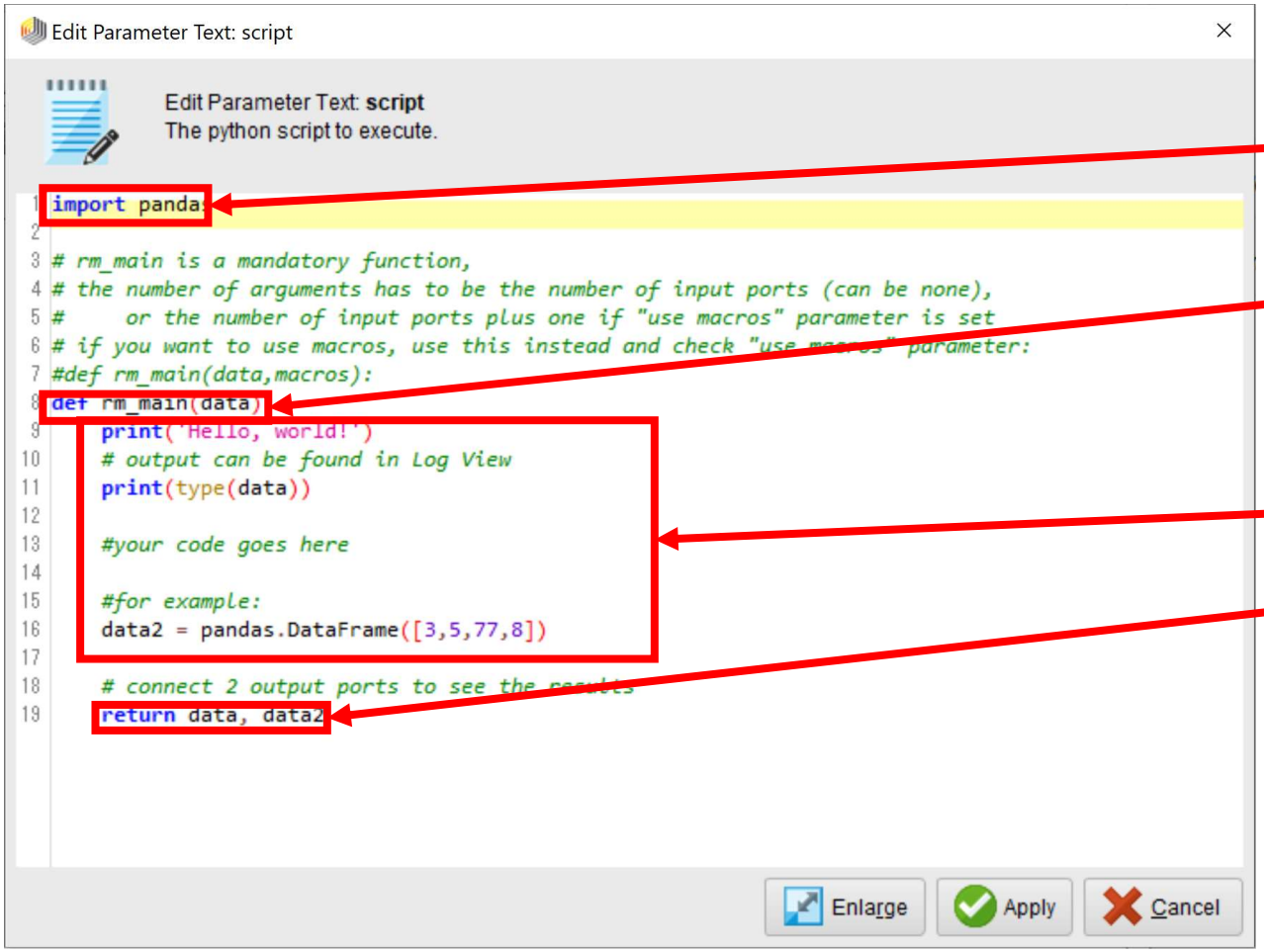
- 「Execute Python/R」の「Show Preview」をクリックして、ポップアップにスクリプトを貼り付けます。
  または
- アクセス可能な場所に保存されているスクリプトファイル （.pyまたはipynb）「Open File」で呼び出します。





2．どのように使用するか

Studioでの利用



- ライブラリをインポートする
- DataTables / Pythonオブジェクトを入れる
- メインコード
- DataTables / Pythonオブジェクトをリターンする

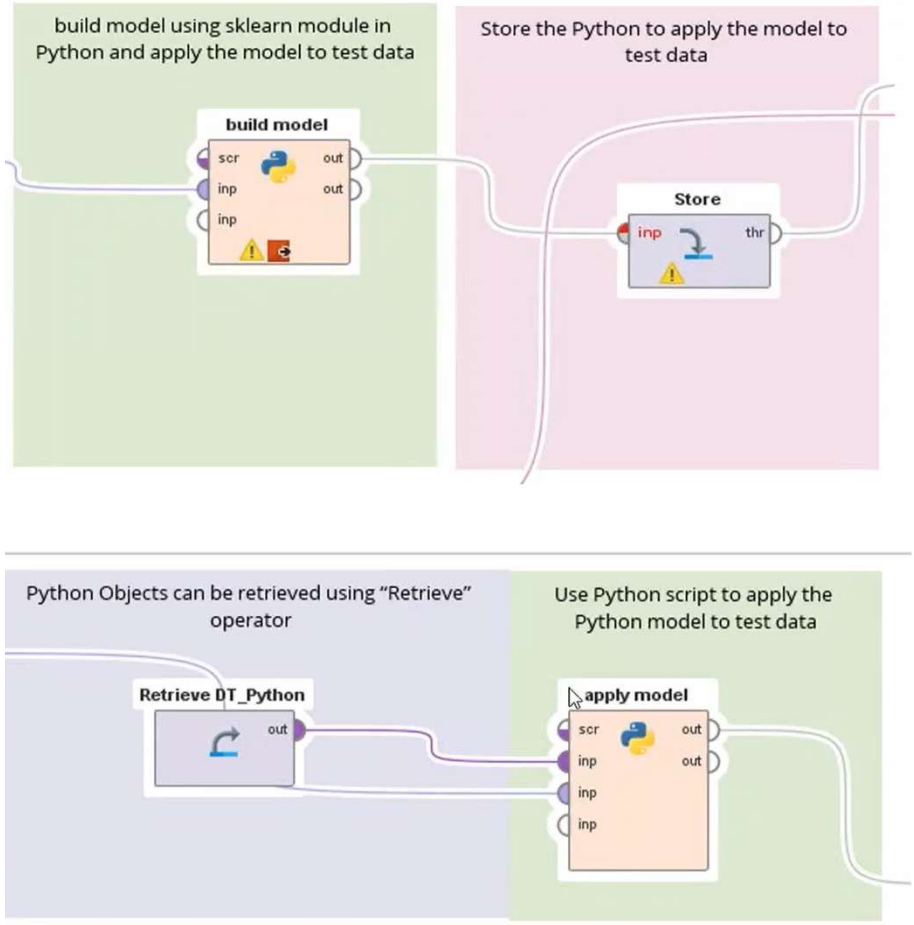
2．どのように使用するか

使用にあたって

- RapidMinerテーブルは自動的にNumpy配列に変換されます。
- Pythonオブジェクトは「Store」オペレータを使用して保存することができます
  例：Pythonで作ったモデルを保存

- Pythonオブジェクトは、「Retrieve」オペレータを使用して取得することができます
  例：以前に作成したモデルを取得

- Pythonオブジェクトは他のPythonスクリプトオペレータに接続できます。
  例：モデルを作成し、別のPythonスクリプトに渡して、独立したテストデータセットで評価する

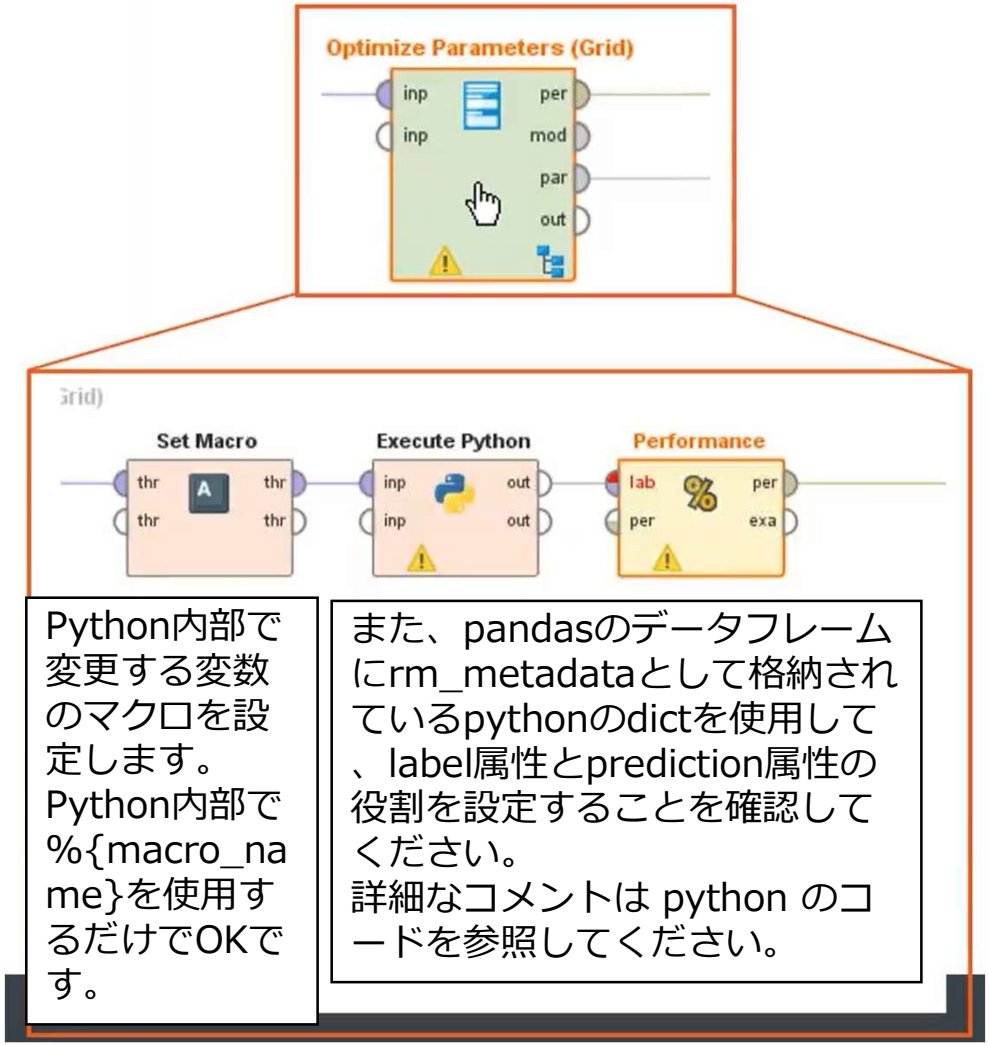


3．Pythonモデルの統合

Pythonモデルの最適化

- Pythonモデルのパラメータは通常のRapidminerモデルと同様に、「Optimize Parameter」オペレータで最適化ができます。

- 「Optimize Patameters」のサブプロセス内に「Set Macro」と「Execute Python」と「Performance」オペレーターを配置します。



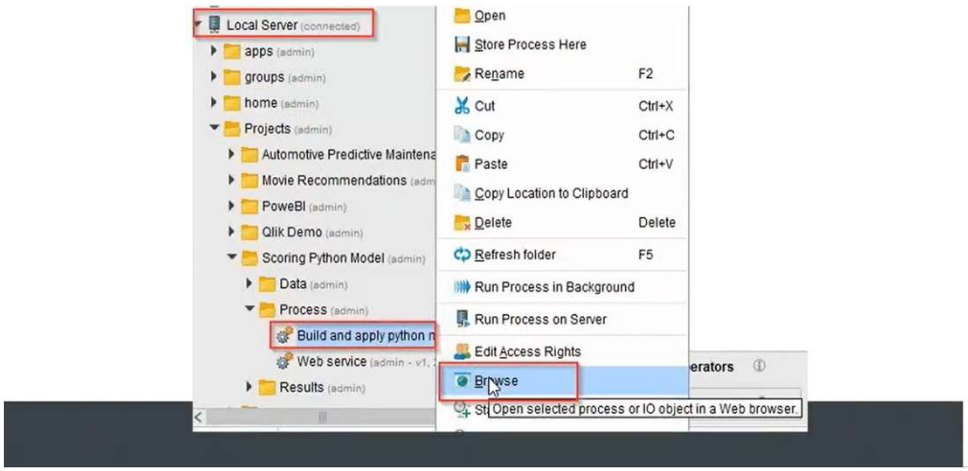
3．Pythonモデルの統合

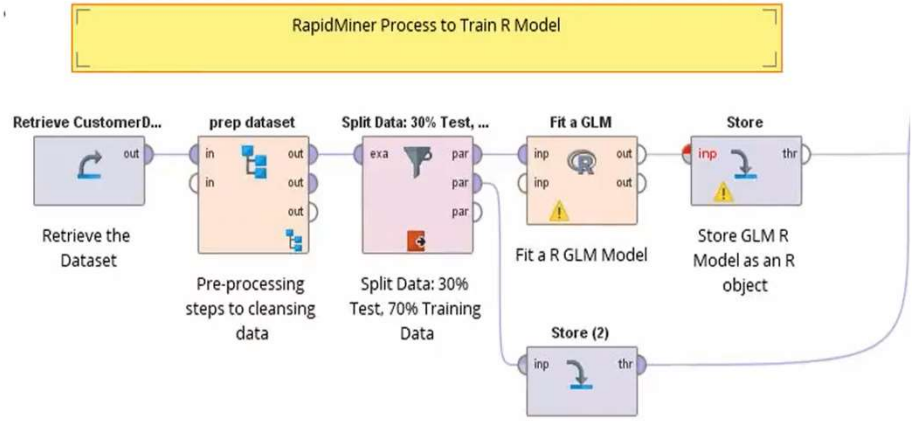
Webサービスとしてデプロイ

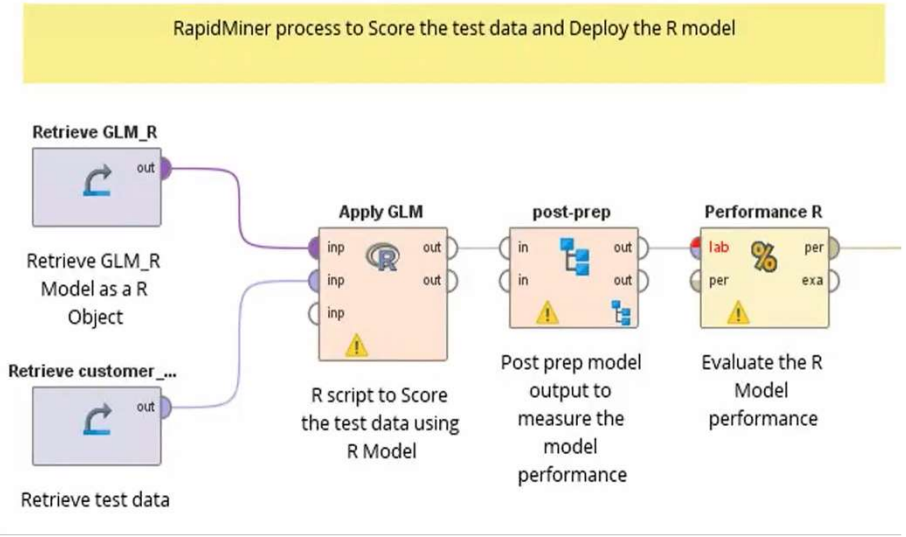
- Python / Rモデルをトレーニングし、モデルにテストデータを適用するプロセスを作成します
- このRapidMinerプロセスをRapidMiner AIHubリポジトリに保存してください

プロセスをWebサービスとして出力する手順

1. ウェブサービスとしてホストされるべきRapidMinerプロセスを選択します。
2. 右クリック→ ブラウズ (Browse)





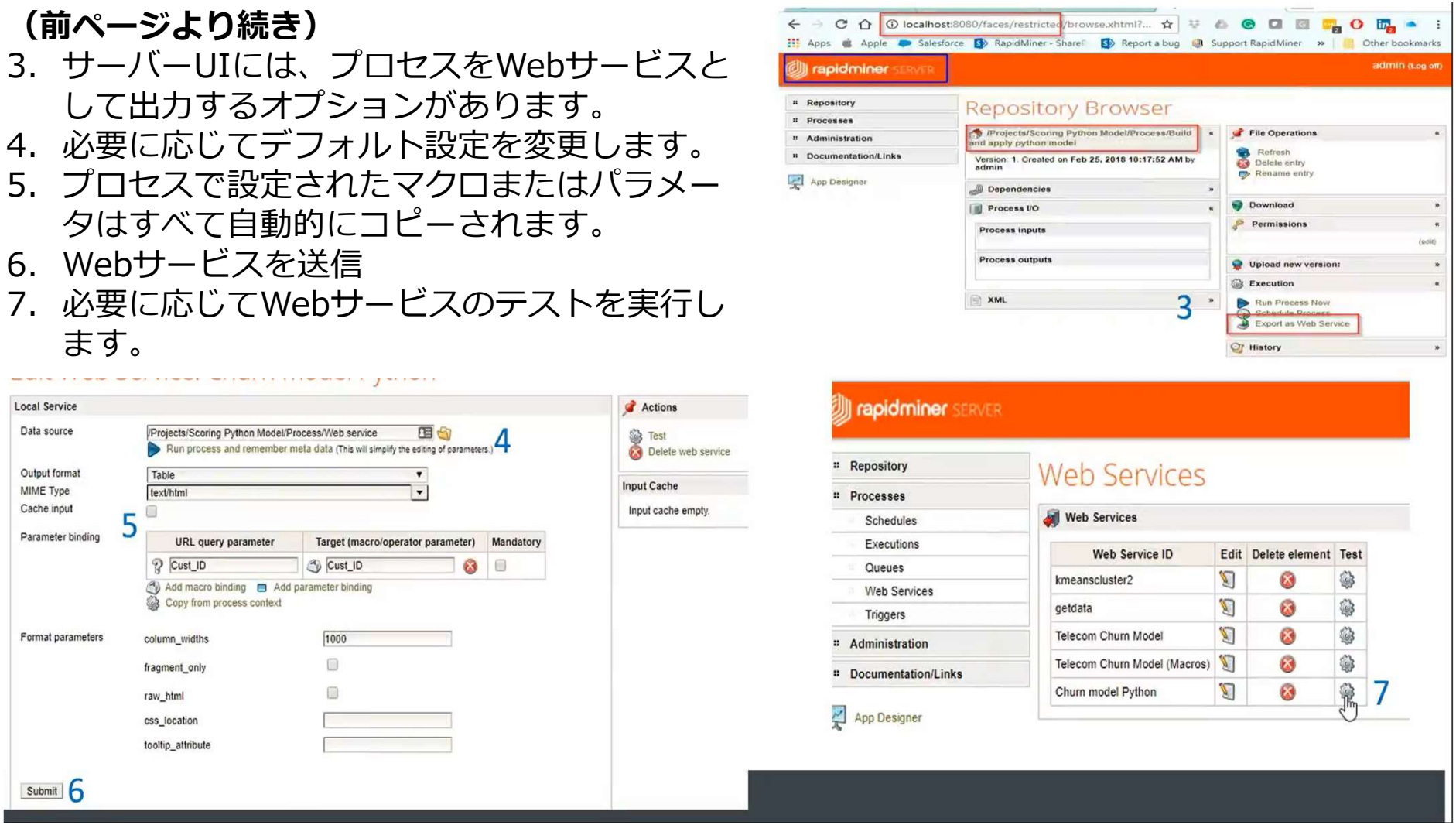


3．Pythonモデルの統合

Webサービスとしてデプロイ

（前ページより続き）

3. サーバーUIには、プロセスをWebサービスとして出力するオプションがあります。
4. 必要に応じてデフォルト設定を変更します。
5. プロセスで設定されたマクロまたはパラメータはすべて自動的にコピーされます。
6. Webサービスを送信
7. 必要に応じてWebサービスのテストを実行します。

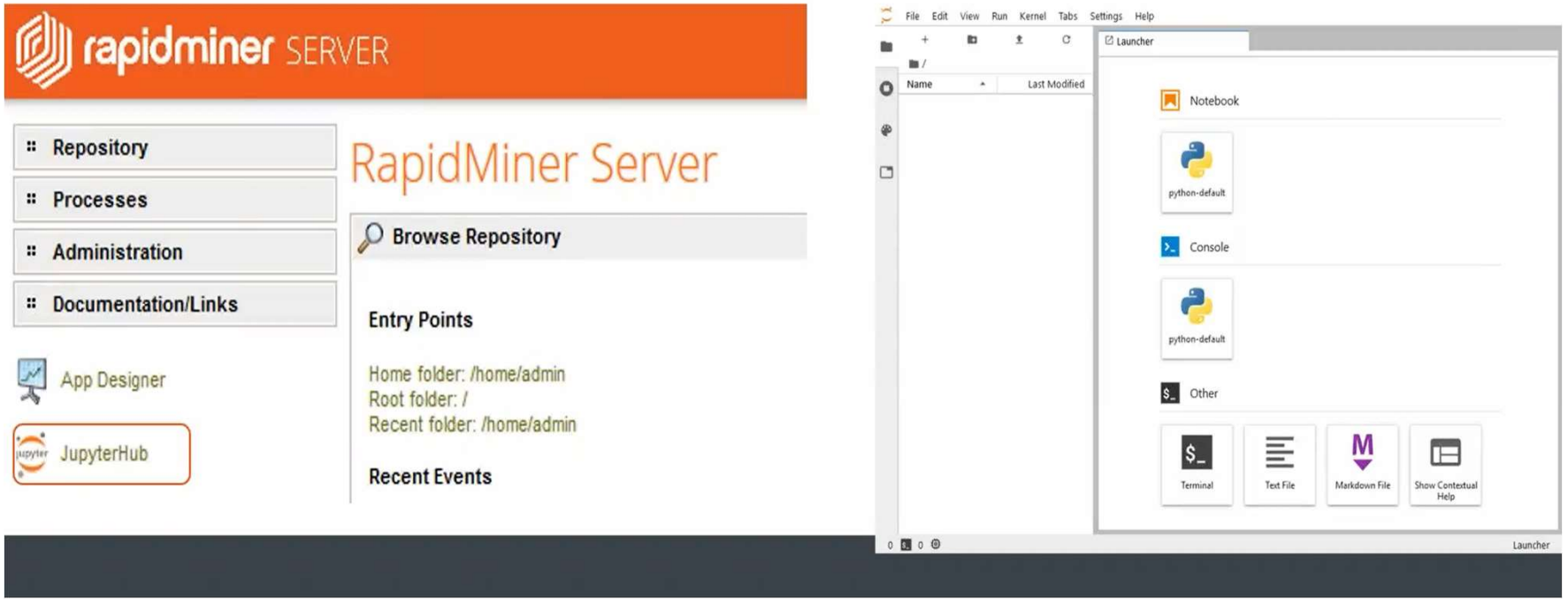


4．RapidMinerとJupyterHubの組み合わせ

JupyterHubとの連携

- JupyterHubはRapidMinerプラットフォームに直接組み込まれるようになりました。
- ユーザーは一元管理および統治された場所でPythonコーディングができます。
- コードベースの作業は視覚的なワークフローベースのプロジェクトおよび自動化されたデータサイエンスと相互運用可能です。

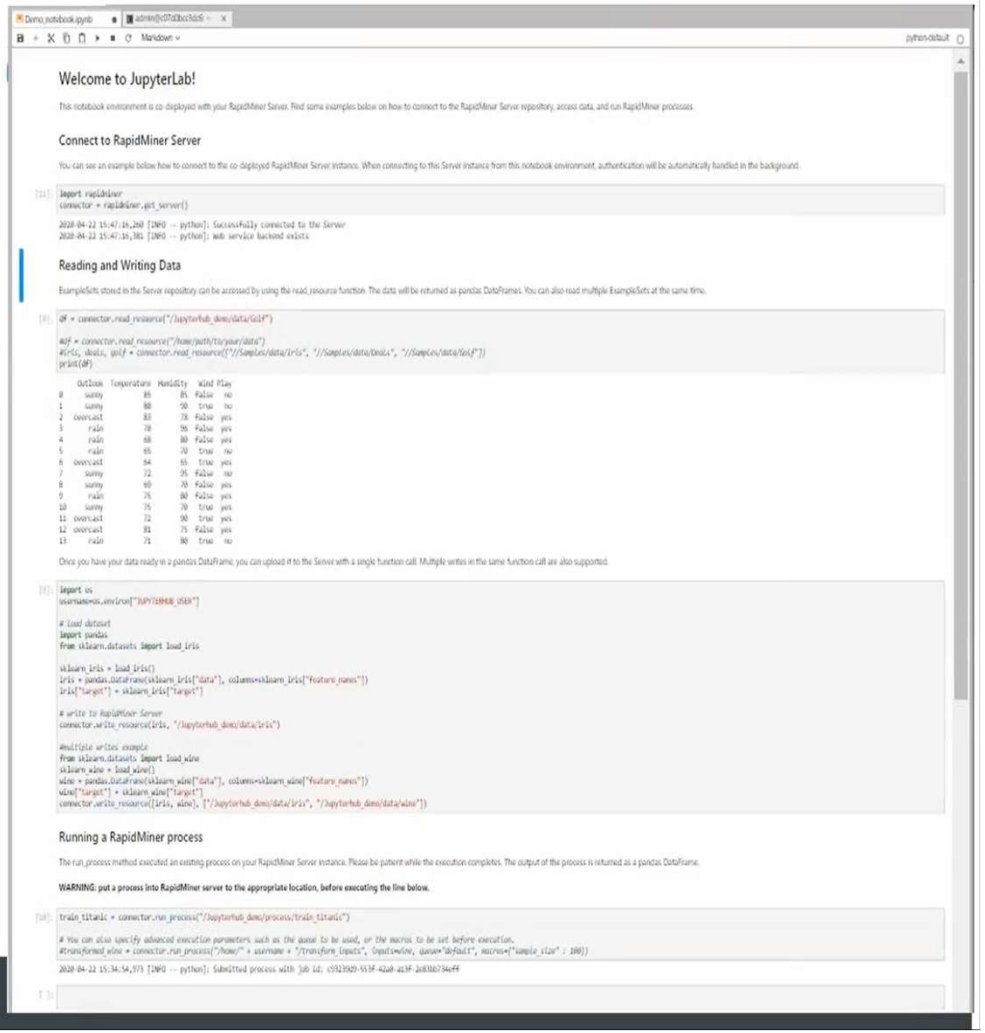
※RapidMiner AIHubの導入が必要です。



4．RapidMinerとJupyterHubの組み合わせ

JupyterHubインターフェースの操作

- RapidMiner AIHubリポジトリのデータ、プロセス、結果などにアクセス出来ます
- JupyterNotebookからRapidMinerワークフローを実行出来ます。
- Notebookからプロセスにリアルタイムの値を渡して、これらのプロセスの実行方法を制御出来ます



4．RapidMinerとJupyterHubの組み合わせ

Python環境の管理

- 簡単にPython環境管理が出来ます

  主な利点:
  - 管理費用の削減
  - 管理エラーのリスクの軽減
  - Pythonコードを含むデプロイされたRapidMinerプロセスの完璧な実行を保証します。