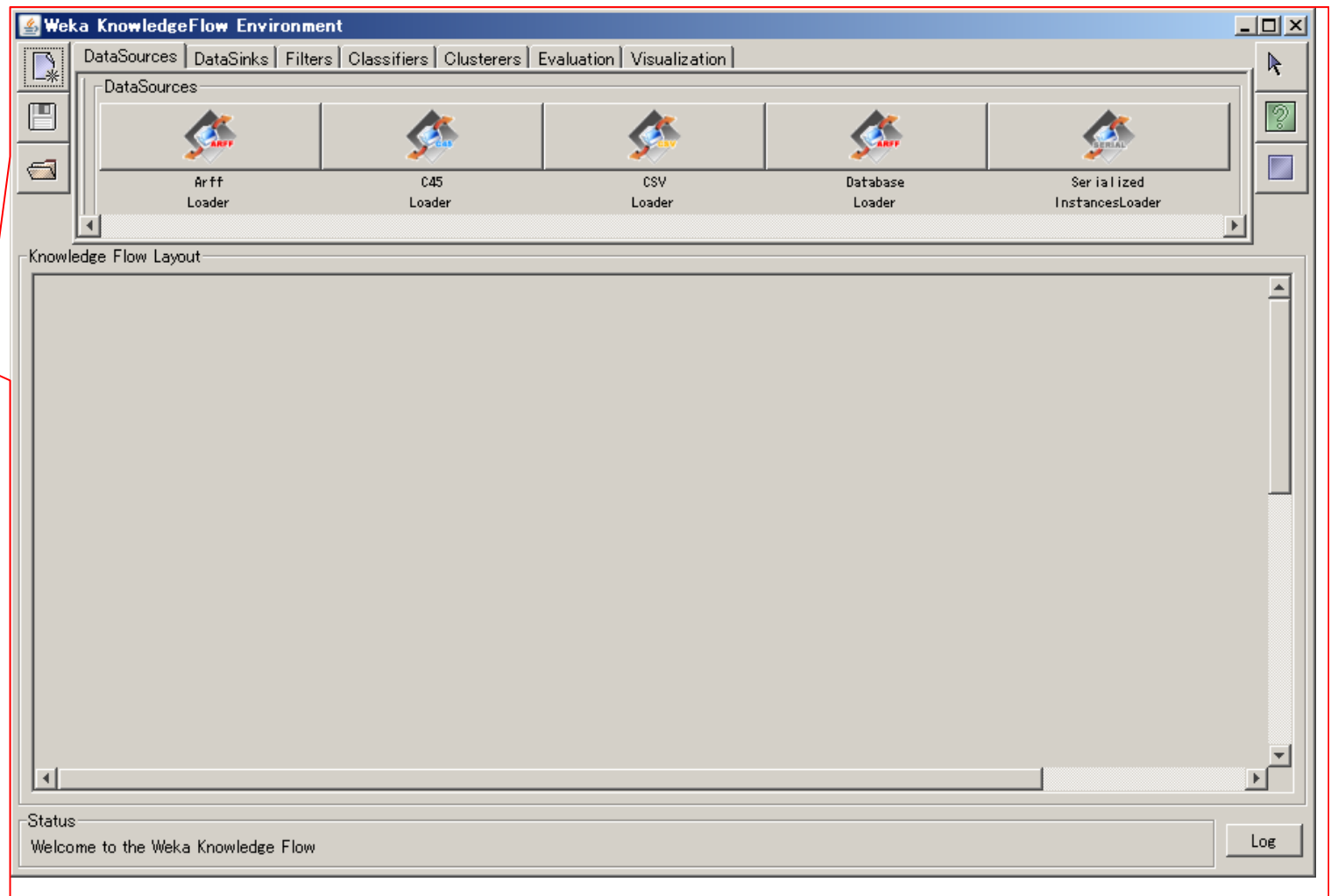




Knowledge Flowの使い方

- Clementine等商用DMソフトウェアがよく使う実行インターフェース
- データの読み込み: DataSources
- データの前処理: Filters
- データの分析: Classifications, Clusterers
- 結果の表示: Visualization, DataSinks
- 精度などの評価法設定: Evaluation

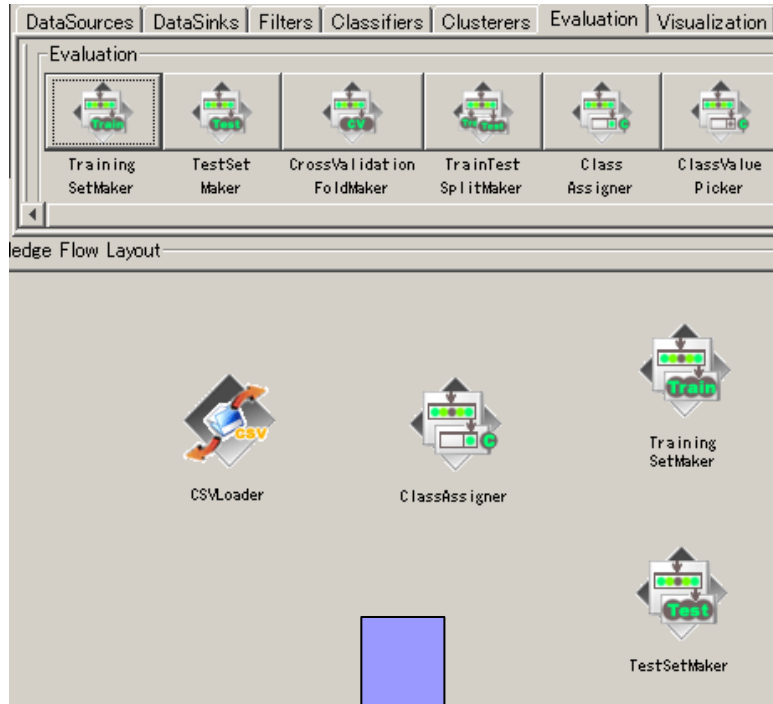
Knowledge Flowの画面



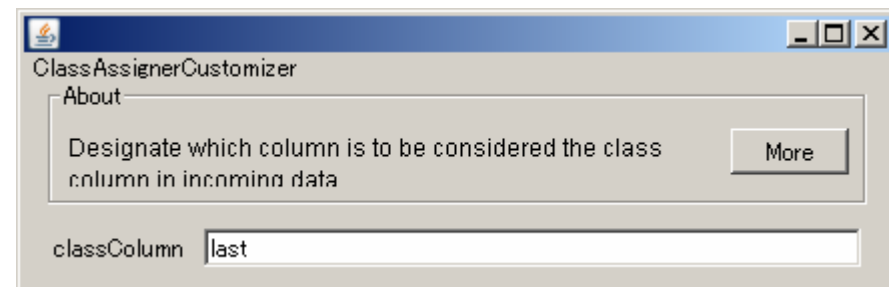
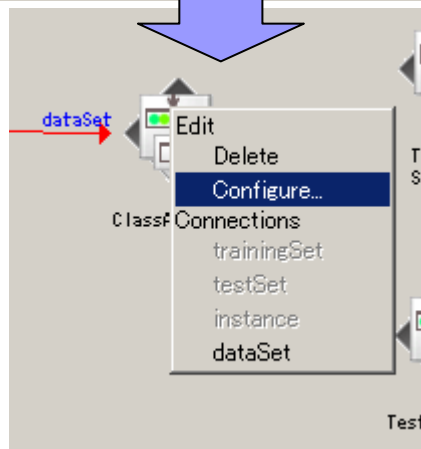
データ(CSVファイル)の読み込み

The image illustrates the process of loading a CSV file into the Weka KnowledgeFlow environment. The main window shows the 'DataSources' panel with 'Arff Loader' and 'C45 Loader' nodes. A 'CSVLoader' node is added to the 'Knowledge Flow Layout'. A context menu is shown over the 'CSVLoader' node with 'Configure...' selected. A file selection dialog is open, showing the file 'diabetes.csv' selected in the 'diabetes.csv' folder. The dialog includes fields for 'ファイル名' (diabetes.csv) and 'ファイルタイプ' (CSV data files), and buttons for '開く' (Open) and '取消し' (Cancel).

評価法の設定(目的変数の割り当て)

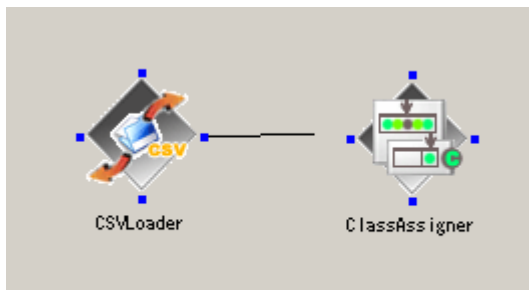
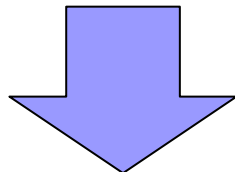
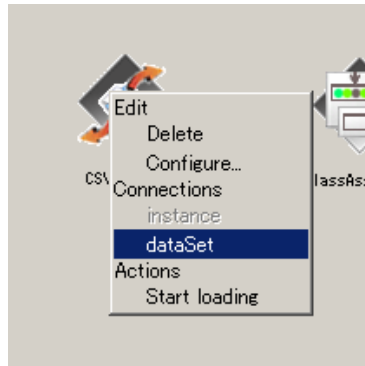


クラス(目的変数)の属性名か
属性番号(最初firstと最後lastも可)を指定

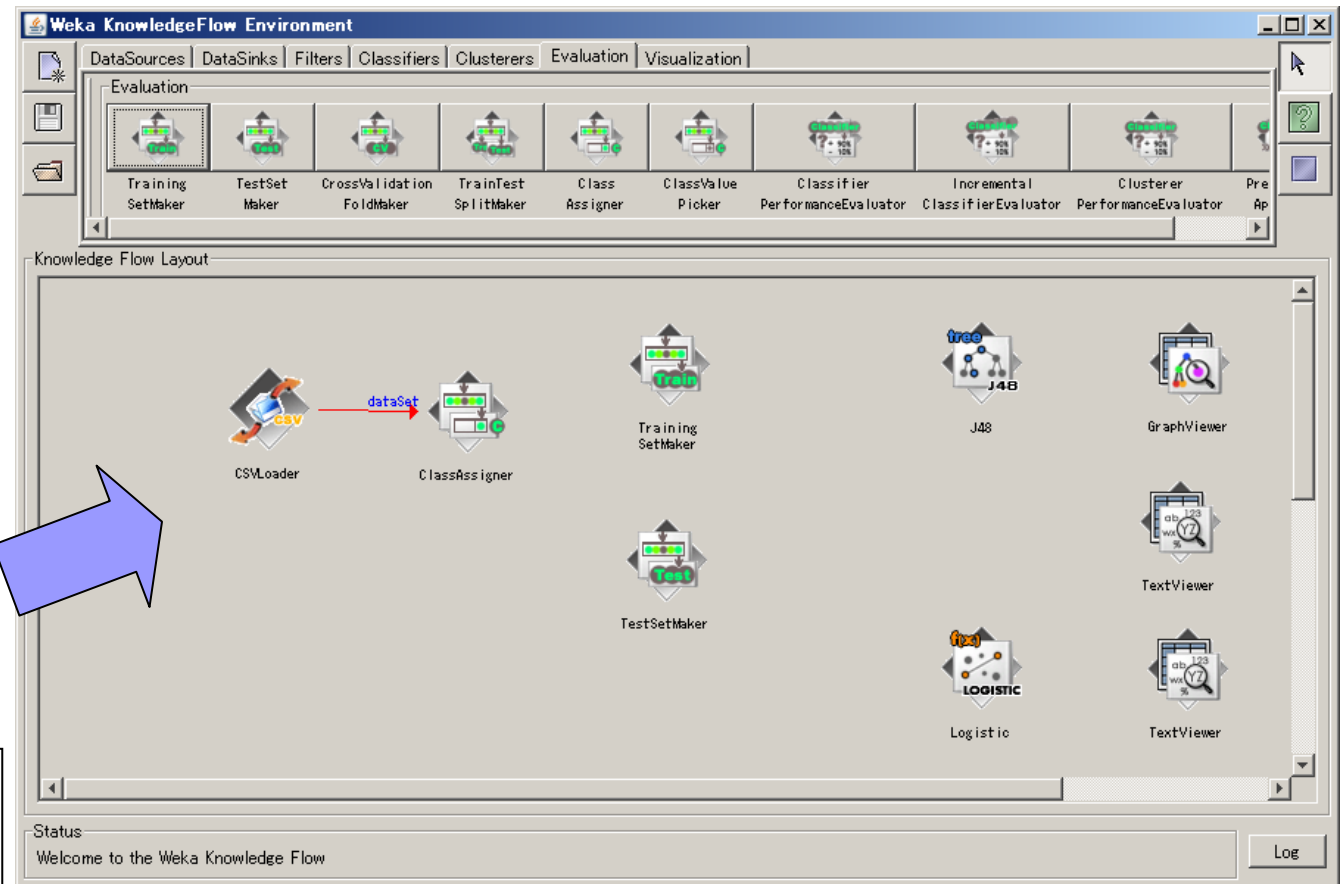


オブジェクト(操作)間の接続

1. オブジェクト上で右クリック



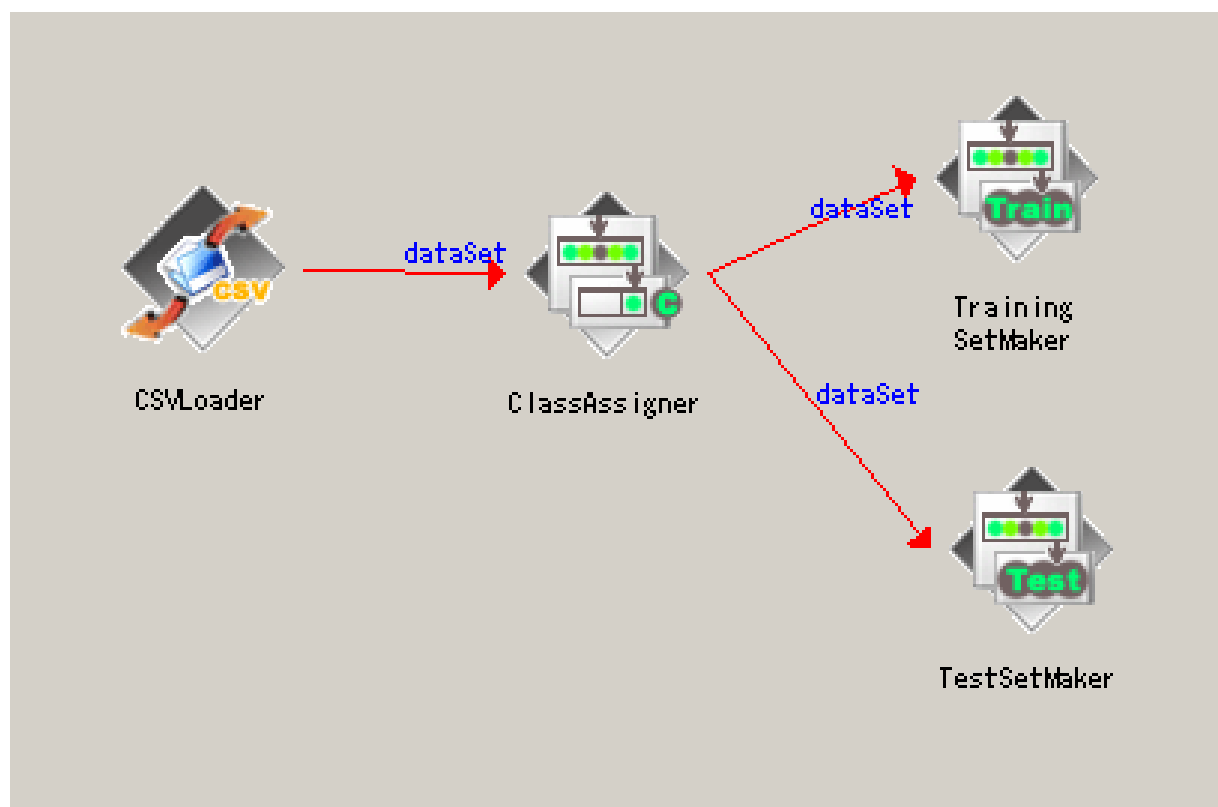
2. 接続先(青■が出る)で右クリック



評価法の設定 (訓練・テストの割り当て)

訓練データ: 学習モデルを生成するためのデータ

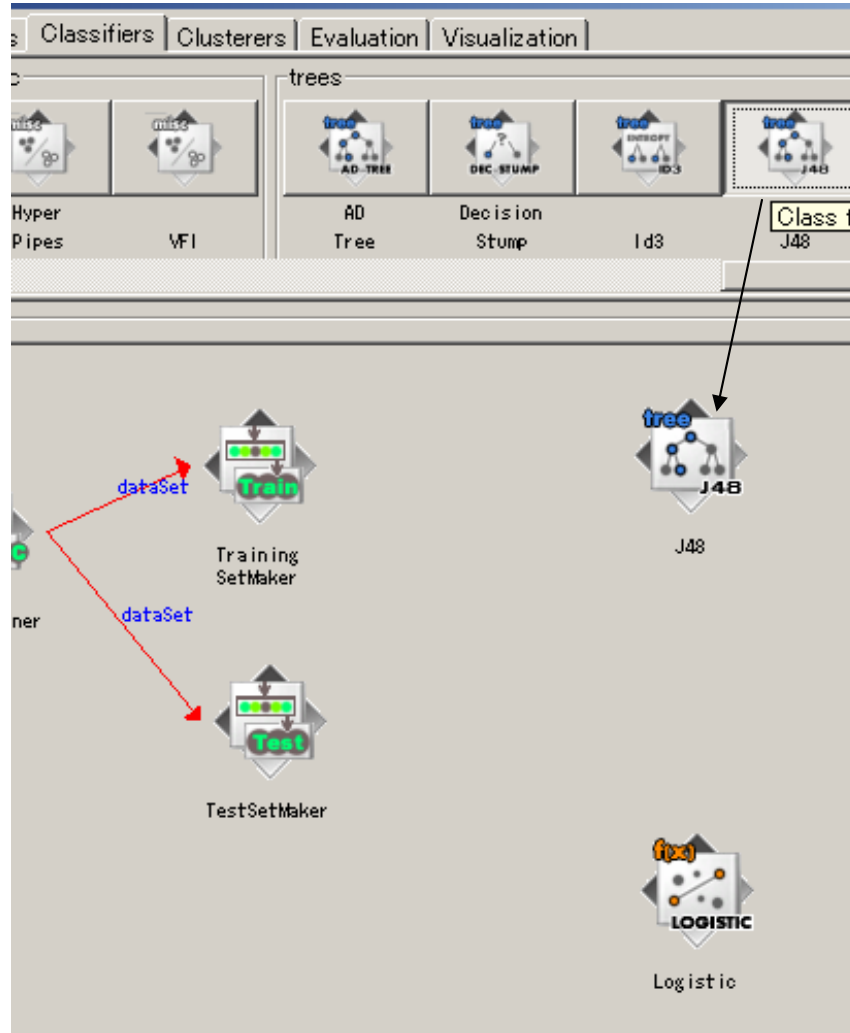
テストデータ: 生成されたモデルを評価するためのデータ. 未知データとも言う



* 上記例の場合は, 同じデータでモデルを作成して評価している

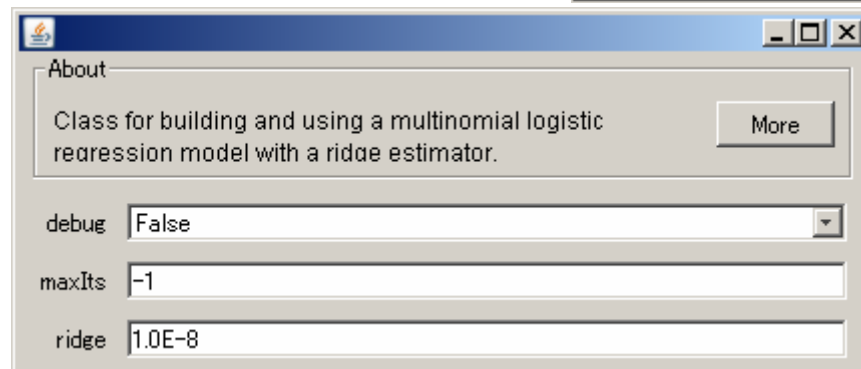
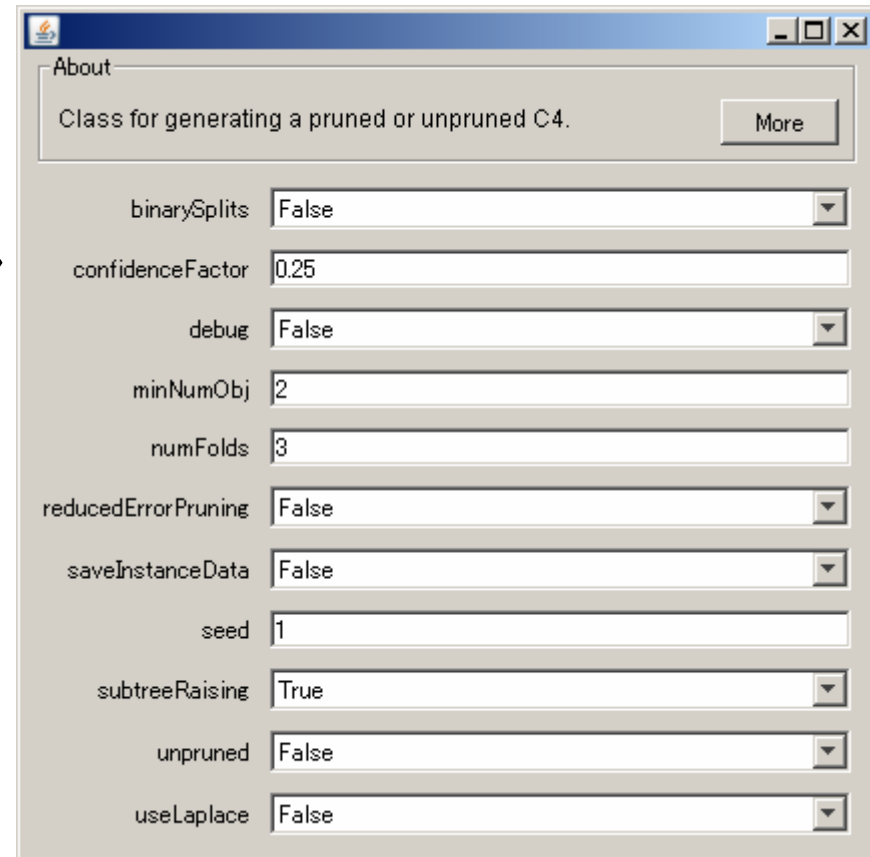
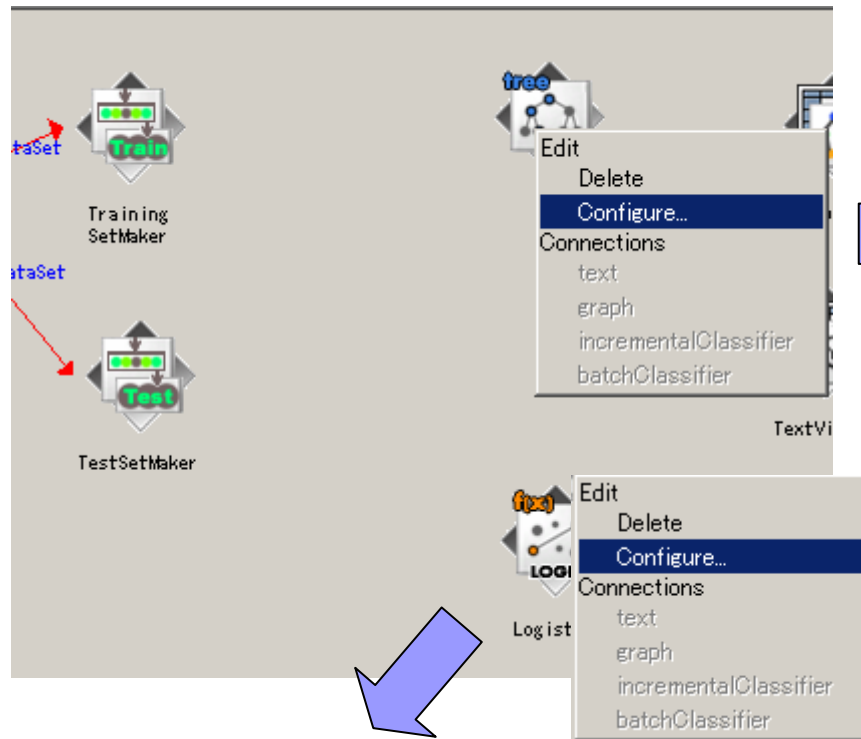
データ分析手法の設定

(決定木生成とロジスティック回帰分析の実行を想定)



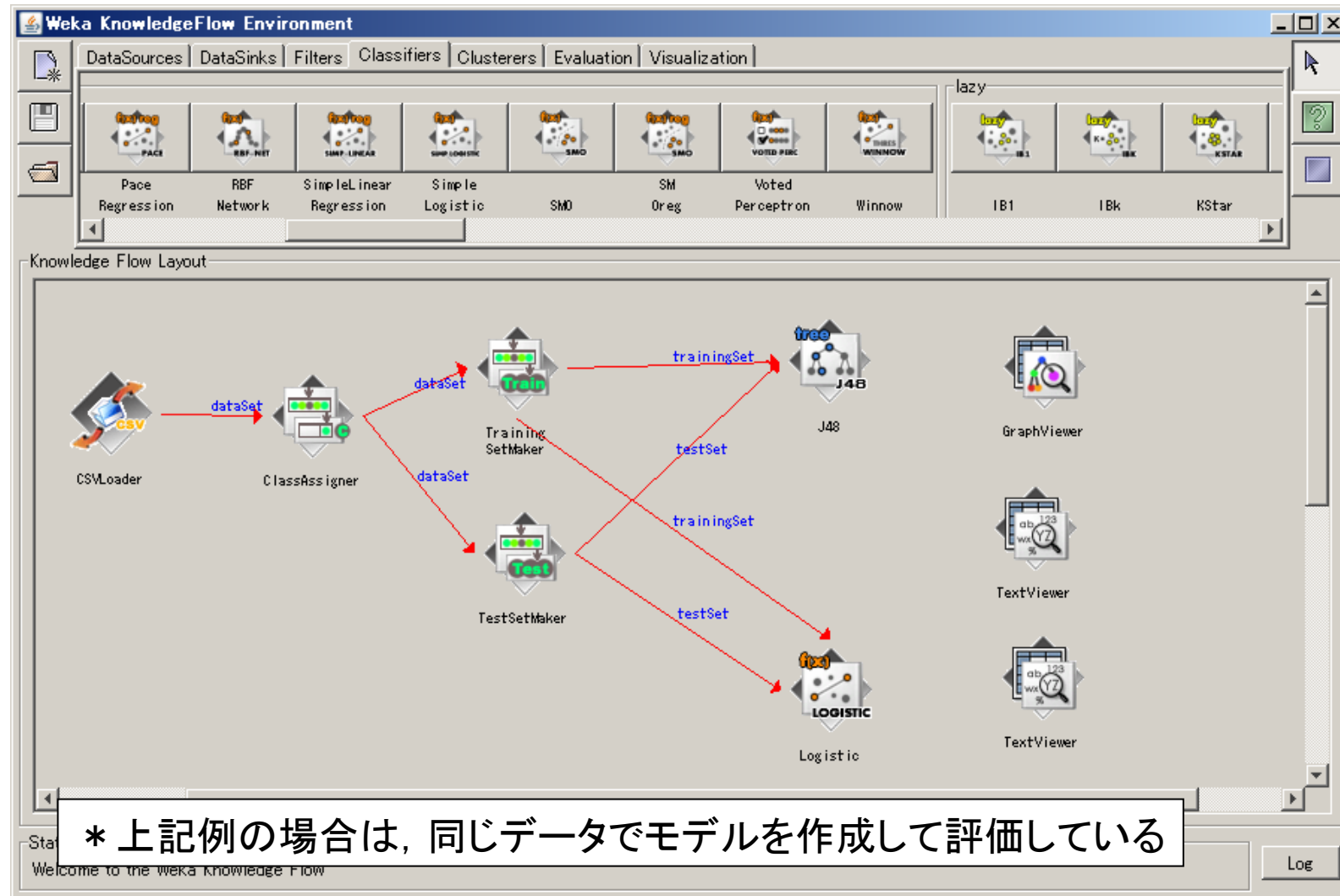
- 決定木・回帰木
 - trees (J4.8(C4.5)など)
- 関数
 - functions (Logistic, linear regression, SVMなど)
- ベイズ則
 - bayes (naïve bayes, bayesian Netなど)
- ルール生成
 - rules
- 事例ベース
 - lazy (k-NNなど)
- メタ学習スキーム
 - meta

データ分析手法の設定

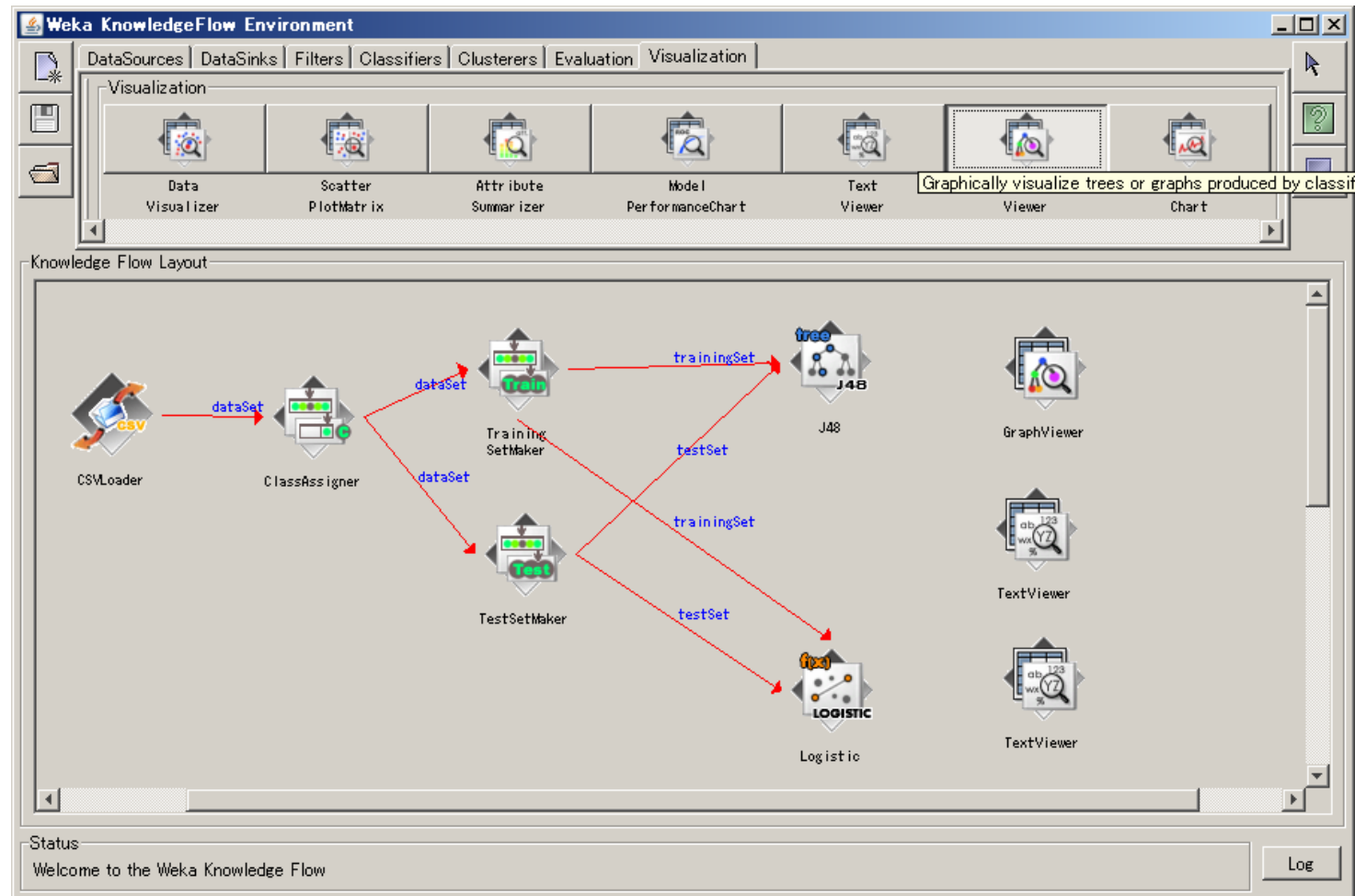


必要に応じて各手法のパラメータを設定

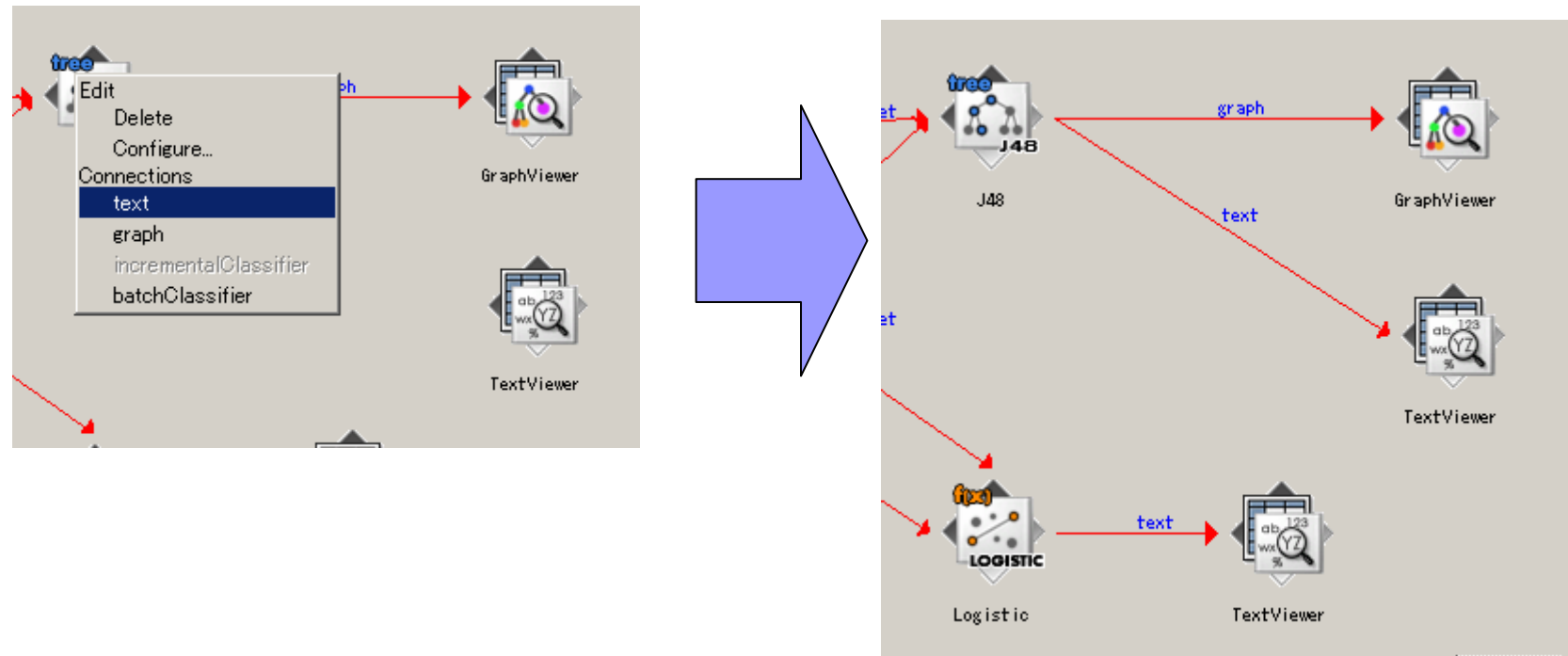
データ分析手法へのデータ入力



結果の出力・視覚化

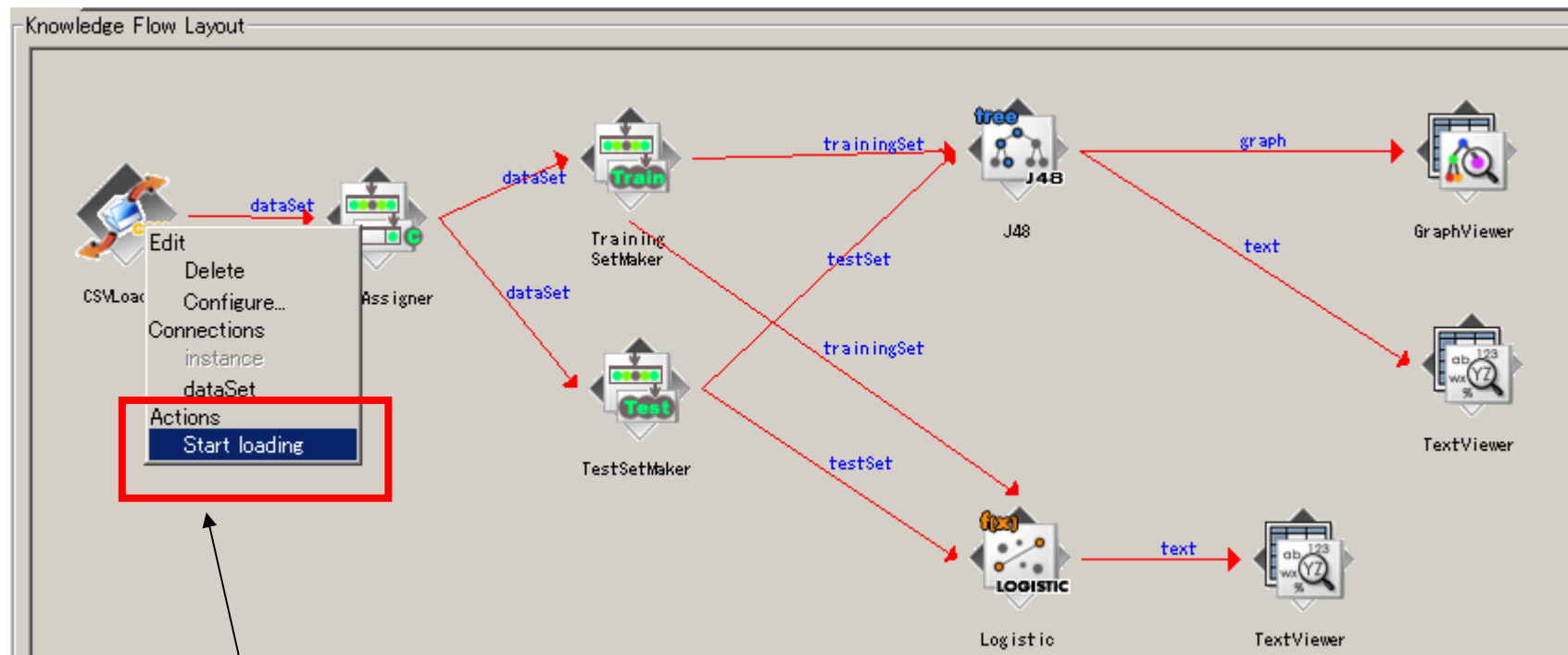


結果の出力のデータフローを作成



ここまででデータフロー図が完成

Knowledge Flowでの実行



Start loadingで実行開始

結果の表示(決定木のグラフ表示)

