Orange Labs

April 2024

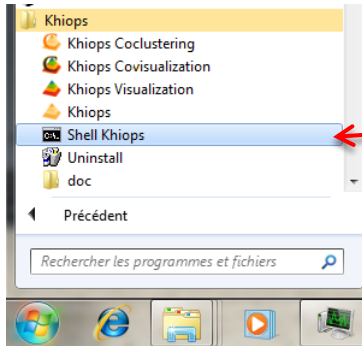# KHIOPS 10.2

KHIOPS SCENARIOS FOR EASY INTEGRATION
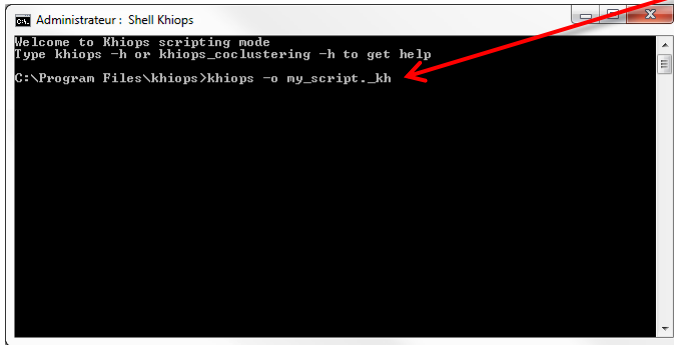
Khiops scenarios

- A Khiops session can be registered in a scenario file, which can be replayed by Khiops in batch mode.

- This allows to automatize data preparation, modeling and deployment in a Data Mining project and to easily integrate the process in any information system.

Recording and replaying a scenario

- Start a Shell Khiops

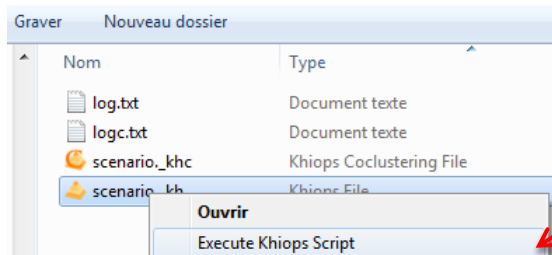- Record a script « automatically» using Khiops user interface

    khiops –o my_script._kh

    o = output

- Replay a script from the shell

    khiops –i my_script._kh

    i = input

- Replay a script from Windows Explorer
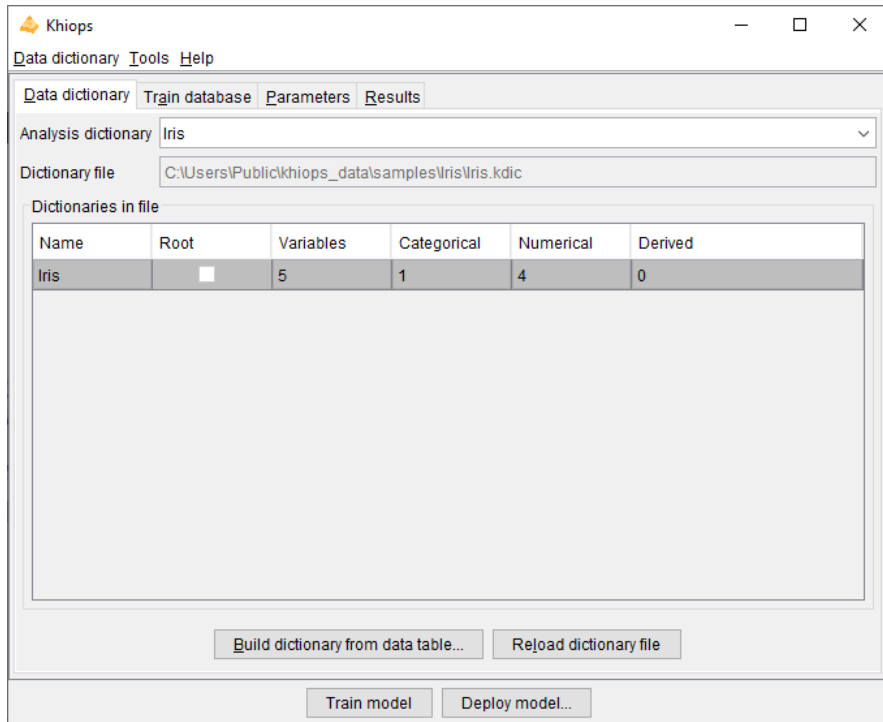
    right click on script file

Recording

- When Khiops is used with option "-o", a scenario is recorded.

    - Khiops –o my_scenario._kh

- Each action on the Khiops user interface is stored in the scenario.

Recording : Example with Iris Dataset

- Open dictionary


```
// -> Khiops
ClassManagement.OpenFile      // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
          // Dictionary file
OK                    // Open
// <- Open
```

Recording : Example with Iris Dataset

- Choose dictionary


```
// -> Khiops
ClassManagement.OpenFile        // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
        // Dictionary file
OK                      // Open
// <- Open
```

Recording : Example with Iris Dataset

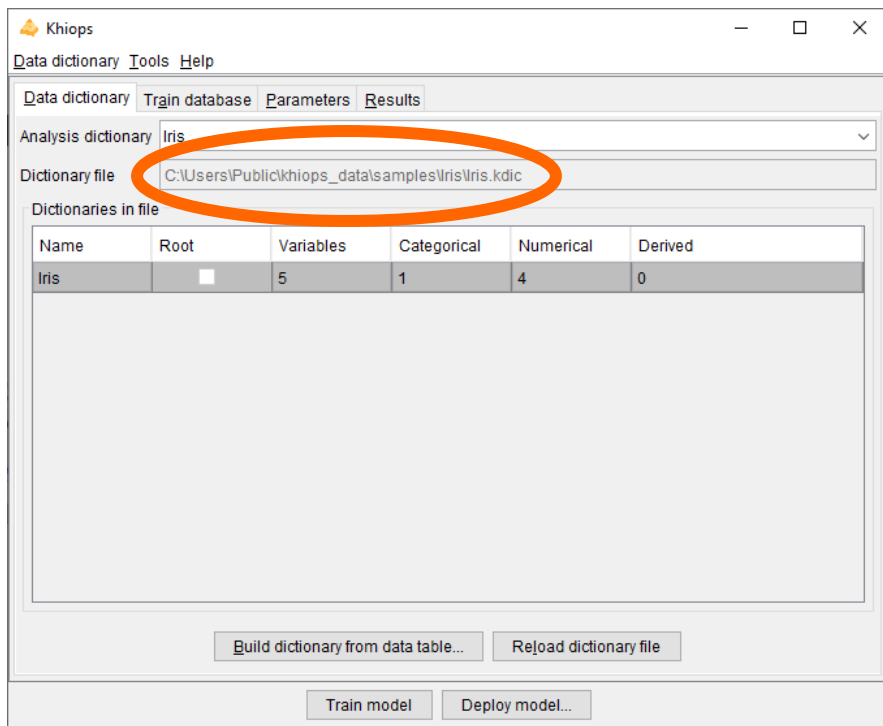- Select data file and split ratio


```
// -> Khiops
ClassManagement.OpenFile      // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
         // Dictionary file
OK                      // Open
// <- Open

TrainDatabase.DatabaseFiles.List.Key Iris       // List item selection
TrainDatabase.DatabaseFiles.DataTableName
         C:\Users\Public\khiops_data\samples\Iris\Iris.txt    // Data table file
TrainDatabase.SampleNumberPercentage 80  // Sample percentage
```

Recording : Example with Iris Dataset

- Select data file and split ratio


```
// -> Khiops
ClassManagement.OpenFile        // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
        // Dictionary file
OK                      // Open
// <- Open

TrainDatabase.DatabaseFiles.List.Key Iris        // List item selection
TrainDatabase.DatabaseFiles.DataTableName
    C:\Users\Public\khiops_data\samples\Iris\Iris.txt    // Data table file
TrainDatabase.SampleNumberPercentage 80  // Sample percentage
```
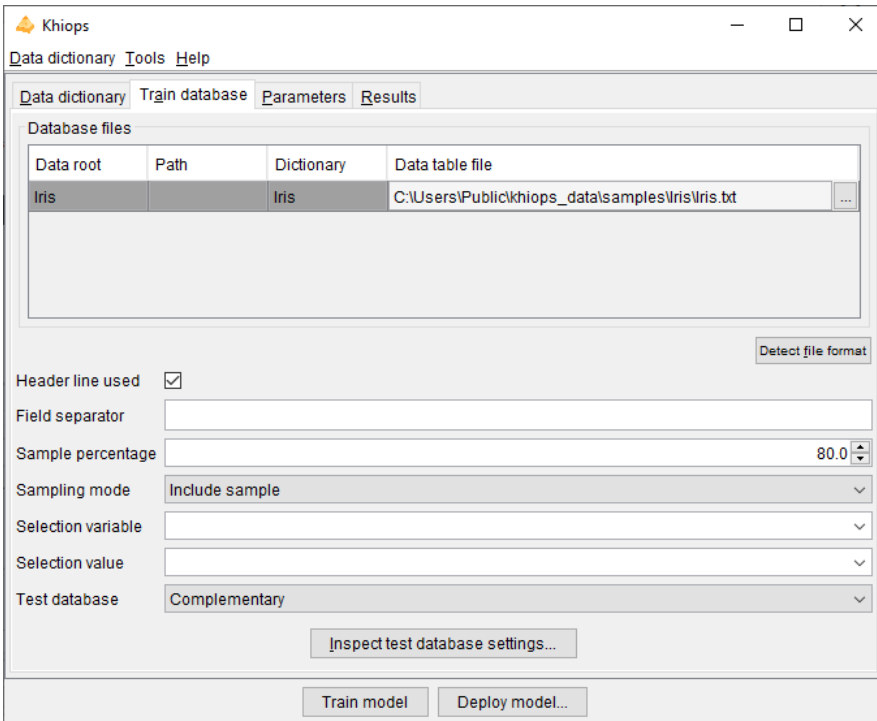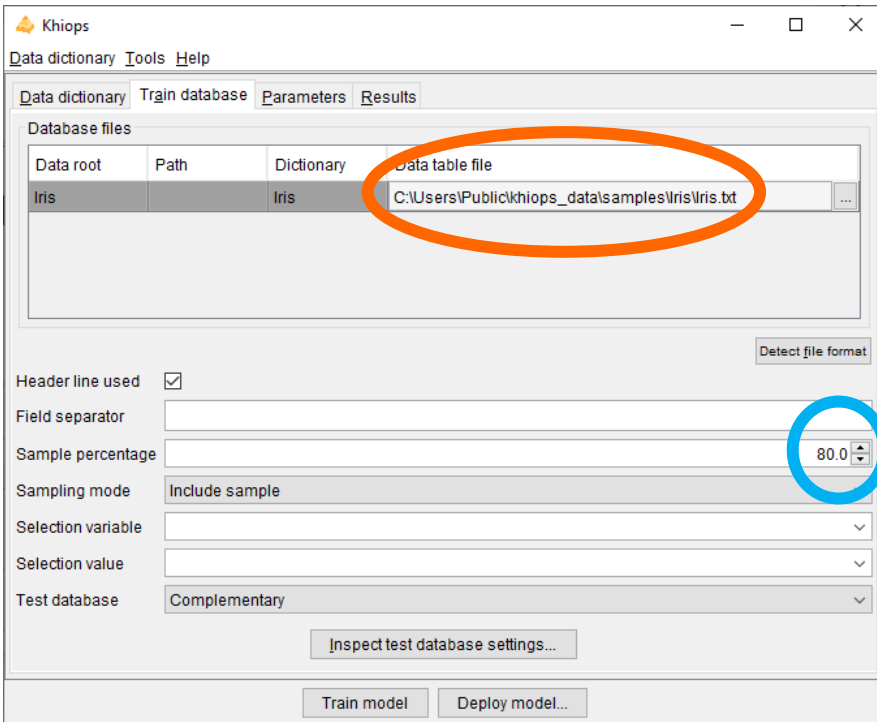
Recording : Example with Iris Dataset

- Select data file and split ratio


```
// -> Khiops
ClassManagement.OpenFile        // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
        // Dictionary file
OK                      // Open
// <- Open


TrainDatabase.DatabaseFiles.List.Key Iris       // List item selection
TrainDatabase.DatabaseFiles.DataTableName
        C:\Users\Public\khiops_data\samples\Iris\Iris.txt    // Data table file
TrainDatabase.SampleNumberPercentage 80  // Sample percentage
AnalysisSpec.TargetAttributeName Class   // Target variable
AnalysisSpec.MainTargetModality Iris-setosa        // Main target value
ComputeStats                // Train model
Exit                    // Close
// <- Khiops


// -> Khiops
OK                      // Yes
// <- Khiops
```
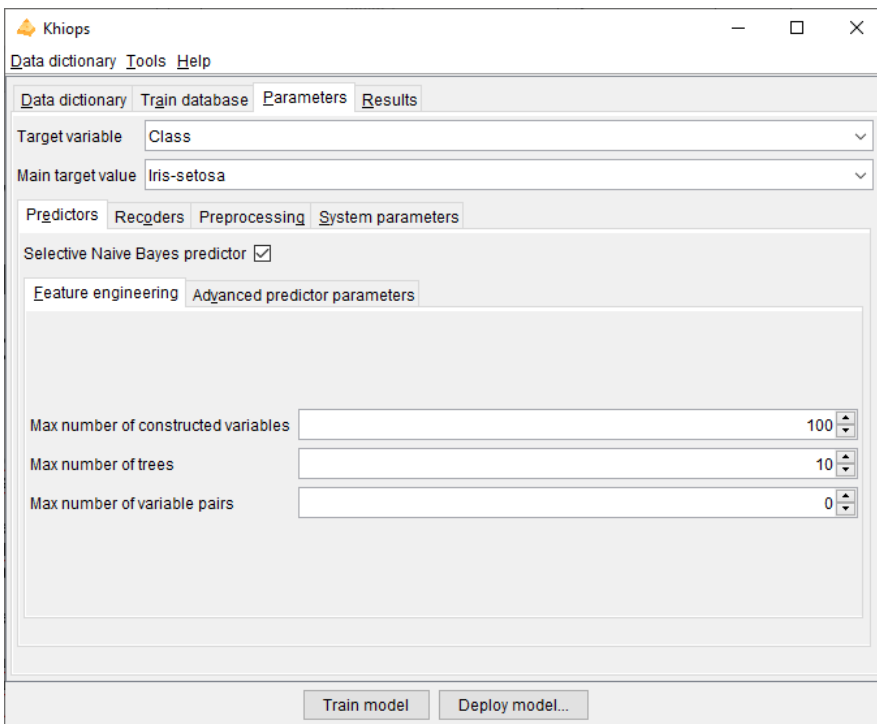
Recording : Example with Iris Dataset
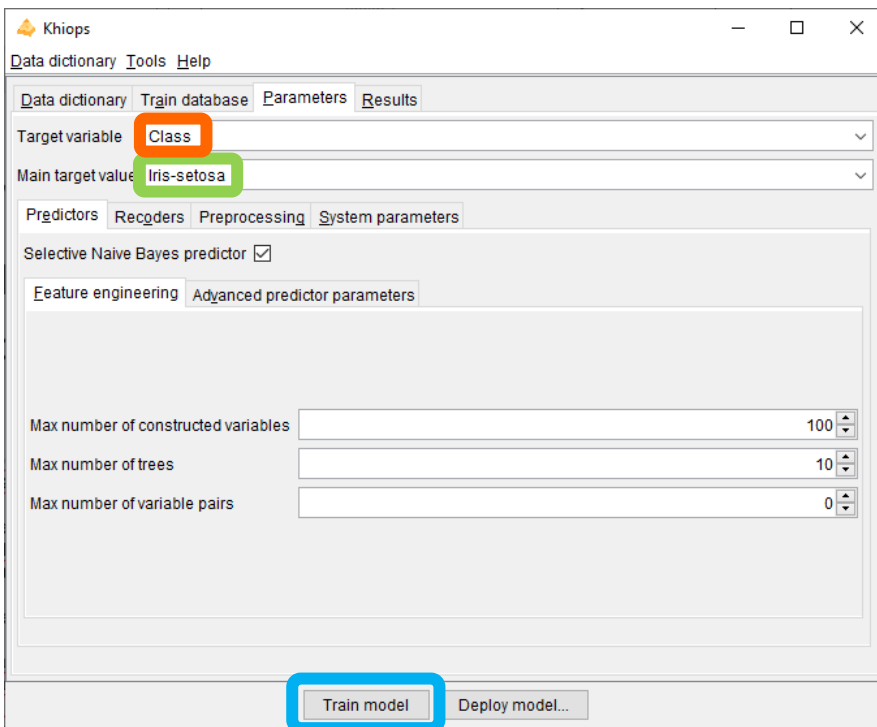
- Select data file and split ratio


```
// -> Khiops
ClassManagement.OpenFile        // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
        // Dictionary file
OK                      // Open
// <- Open

TrainDatabase.DatabaseFiles.List.Key Iris        // List item selection
TrainDatabase.DatabaseFiles.DataTableName
        C:\Users\Public\khiops_data\samples\Iris\Iris.txt    // Data table file
TrainDatabase.SampleNumberPercentage 80  // Sample percentage
AnalysisSpec.TargetAttributeName Class   // Target variable
AnalysisSpec.MainTargetModality Iris-setosa        // Main target value
ComputeStats                // Train model
Exit                    // Close
// <- Khiops


// -> Khiops
OK                      // Yes
// <- Khiops
```

Playing a scenario

- When Khiops is used with option "-i", the scenario is replayed.

  - khiops –i my_scenario._kh

Dealing with scenarios

- You can edit scenarios in a text editor to apply it on another dataset.

Dealing with scenarios

- You can edit scenarios in a text editor to apply it on another dataset.

```
// -> Khiops
ClassManagement.OpenFile      // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Iris\Iris.kdic
        // Dictionary file
OK                    // Open
// <- Open

TrainDatabase.DatabaseFiles.List.Key Iris
TrainDatabase.DatabaseFiles.DataTableName
        C:\Users\Public\khiops_data\samples\Iris\Iris.txt
TrainDatabase.SampleNumberPercentage 80  // Sample percentage
AnalysisSpec.TargetAttributeName Class   // Target variable
AnalysisSpec.MainTargetModality Iris-setosa      // Main target value
ComputeStats            // Train model
Exit              // Close
// <- Khiops


// -> Khiops
OK                    // Yes
// <- Khiops
```

```
// -> Khiops
ClassManagement.OpenFile      // Open...

// -> Open
ClassFileName C:\Users\Public\khiops_data\samples\Adult\Adult.kdic
        // Dictionary file
OK                    // Open
// <- Open

TrainDatabase.DatabaseFiles.List.Key Adult
TrainDatabase.DatabaseFiles.DataTableName
        C:\Users\Public\khiops_data\samples\Adult\Adult.txt
TrainDatabase.SampleNumberPercentage 80  // Sample percentage
AnalysisSpec.TargetAttributeName class   // Target variable
AnalysisSpec.MainTargetModality more       // Main target value
ComputeStats            // Train model
Exit              // Close
// <- Khiops


// -> Khiops
OK                    // Yes
// <- Khiops
```

Dealing with scenarios

- You can replace any element of a scenario with the « -r » option

  - khiops –i my_scenario._kh –r to_replace:new

- To run analysis with 90% of instances in train instead of 80%

  - khiops –i my_scenario._kh –r 80:90

Dealing with scenarios

- We can easily make the scenario more generic

```
// -> Khiops
ClassManagement.OpenFile      // Open...

// -> Open
ClassFileName $DICTIONARY_FILE$     // Dictionary file
OK                    // Open
// <- Open
TrainDatabase.DatabaseFiles.List.Key $DICTIONARY_NAME$
TrainDatabase.DatabaseFiles.DataTableName $DATA$

TrainDatabase.SampleNumberPercentage $TRAIN_PERCENTAGE$
AnalysisSpec.TargetAttributeName $TARGET_NAME$
AnalysisSpec.MainTargetModality $TARGET_MODALITY$
ComputeStats              // Analyse database
Exit                  // Close
// <- Khiops


// -> Khiops
OK                    // Close
// <- Khiops
```

Dealing with scenarios

- Beware of ambiguities :
    - replace DATA
    - replace DATA_PATH

- Recommendations :
    - $DATA$
    - $DATA_PATH$

Tips and tricks

- A scenario is automatically generated by khiops in the directory *C:\Users\<username>\khiops_data\lastrun*

- You don't know the syntax and you want to add features to your scenario ?
  - Just click on the khiops buttons and open the scenario in the lastrun directory

- Use the « -b » option in conjunction with « -i » and « -r » to replay scenarios silently (without a user interface)
- Use the « -e <file> » to store the results logs in a file

Integration with other programming languages

- If you need to start a Khiops process from your favorite programing language: C++, Java, Java script, MATLAB, R…

  - Record a scenario using Khiops application

  - Make the scenario more generic

  - Prepare a Khiops command line with options -i, -r, -b, -e

  - Call Khiops with this command line and the generic scenario from your favorite language

    - Example
      - C++: system(command)
      - Java: Process process = Runtime.getRuntime().exec(command);
      - …

Note on backwards compatibility

- Khiops scenario are not backwards compatible

- In the event of a new version of Khiops
  - Simply re-register a scenario and make it generic
  - Reuse the same integration process by just updating the scenario files