Introduction

This three-day course offers in-depth training on the fundamentals of Advanced Data Mining with the KNIME® environment. The training is aimed at people who are already familiar with KNIME® and have a first prior experience in data mining. The main theme covered throughout the course is the implementation of more complex workflows by learning more advanced programming concepts such as the encapsulation of nodes in components, the use of variables, the notion of streaming, the implementation of optional connections such as CASE or IF-THEN-ELSE and the use of different loop variants in the KNIME® platform. The training covers in a second time, the learning of advanced data mining nodes for more complex data transformation such as the creation and management of collections, the grouping (Group By) and ungrouping (Ungroup), the transposition (Transpose), the pivoting (Pivoting) and the unpivoting (Unpivoting) of data, among others.

Who is this course for?

This course is intended for scientists who want to acquire a deeper knowledge of data mining on the KNIME® environment with many practical examples to illustrate the theory.

To follow this course, you should . . .

•Already have some experience using the KNIME® environment equivalent to the level reached during the PIKAÏROS course « An Introduction to Data Mining with KNIME® » or have obtained a KNIME® L1 certification.

•Understand basic computer English (the KNIME® interface is in English).

Terms, access time and adaptation to people with disabilities

This course is entirely based on the KNIME educational software with the implementation of workflows as a learning method.

A continuous evaluation by MCQ is carried out on the exercises proposed during the training.

This course is offered only face-to-face in intra-company at the client's premises.

Possibility of adapting the time and mode of training to people with disabilities.

Consider 1 month waiting time to access the training from the order.

Full programme

First day

Introduction to advanced programming in the KNIME® environment

Meta-Nodes and Components

- Definition and utility of meta-nodes and components
- External configuration of components
- Programming inside the components
- Definition, configuration and execution of nodes in streaming mode

Using variables in KNIME® Protocols

- Definition and utility of variables
- Parameterization of nodes by variables
- Initialization of variables by Quickforms
- Programmatic initialization of variables, modification, data transfer between variables and tables, and vice versa
- Discovery of the most useful nodes for the use of variables in a protocol

Flow control in the programming of a protocol

- Definition and utility of flow control, CASE or IF Switches
- Discovery of the most useful nodes for flow control in a protocol

Exercise: Reorganizing a complex workflow using variables and meta-nodes

- Installation, study and understanding of a complex workflow with reading, transformation and writing of cheminformatics data.
- Transformation of the workflow with parameterization by variables.

Second day

Programming using loops

Notion of loop and elementary variants

- Definition and usefulness of loops
- Context to promote or to avoid when using loops in the KNIME® environment
- Elementary loops in KNIME® (Count Loop, Generic Loop)

Iteration-specific loops on rows and columns

- Loops for iteration on the rows of a table
- *Row to Variable, Chunk Loop, Interval Loop, Group Loop*
- Loops for iteration on the columns of a table
- *Column List Loop*
- Loops with generation of variables by iteration
- *Row To Variable*

Advanced use of loops in the KNIME® environment

- Recursive loops
- Loops with breakpoint
- Loops with control and error correction (Try & Catch)

Exercise 1: Programming loops in a data cheminformatics context

- Generation of fingerprints and physicochemical properties.
- Implementation of a simple predictive model on an ADMET or therapeutic target.
- Training, testing and optimization of a predictive model using loops.
- Selection of homogeneous portions in the workflow and encapsulation of these portions into components to improve the modularity of the workflow, its readability and facilitate the reuse of sub-workflows.

Third day

Discovery of advanced nodes for protocol optimization

Extraction and reuse of protocol meta-information

- KNIME® environment information
- Control of the execution time of KNIME® protocols
- Generic information on the tables

Advanced operations with tables

- Grouping and ungrouping of data
- Transposition of data
- Pivoting and unpivoting of data
- Junction of tables
- Convert multiple columns to nominal column and vice versa
- Manipulation of collections: creation and division of collections

Advanced data manipulation

- Creation of bit vectors (Create Bit Vector)
- Extraction of bit vectors (Expand Bit Vector)

Discovery of new KNIME® nodes of general utility

- **Read nodes:** List Files, Read Images, Table Reader, Model Reader, Excel Reader
- **Write nodes:** Write Images, Table Writer, Model Writer, Excel Writer & Appender
- **Generic nodes:** RowID and Constant value Column

Exercise: Reading, transforming and saving data from different cheminformatics sources using the nodes studied during the training.