Data Mining

SPSS Clementine 12.0

3. Projects

Fall 2009

Instructor: Dr. Masoud Yaghini

Outline

- **Introduction**

- **Building a Project**

- **References**

Introduction

Introduction

- A project is a group of files related to a data mining task.

- Projects include
  - data streams,
  - graphs,
  - generated models,
  - reports, and
  - anything else that you have created in Clementine.

- Note:
  - If the Projects tool is not visible in the Clementine window, choose Project from the View menu.

Clementine

Introduction

- Using projects, you can:
  - Annotate each object in the project file.
  - Use the CRISP-DM methodology to guide your data mining efforts.
  - Add non-Clementine objects to the project, such as a PowerPoint slide show used to present your data mining goals or white papers on the algorithms that you plan to use.
  - Produce both comprehensive and simple update reports based on your annotations.

Introduction

- Objects that you add to a project can be viewed in two ways: Classes view and CRISP-DM view.
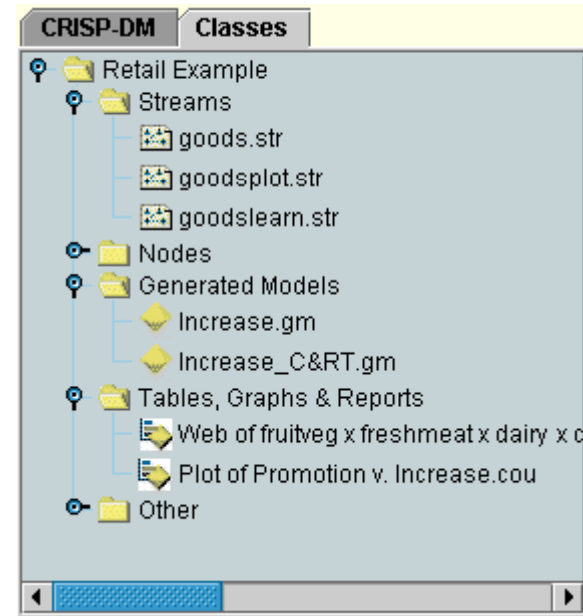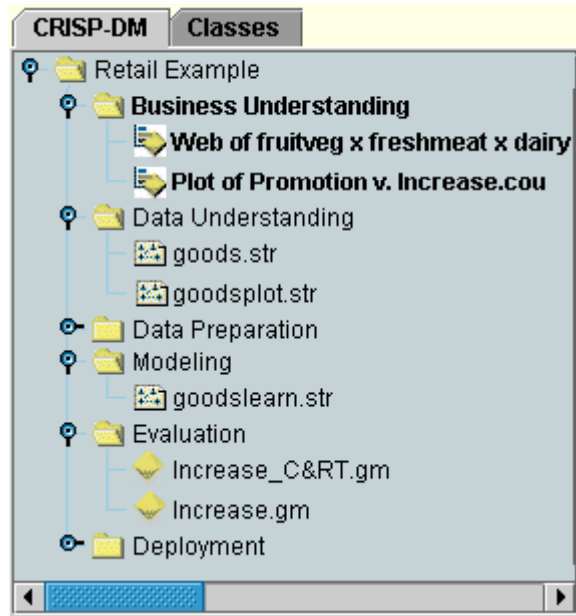
Introduction

- **Setting the Default Project Phase**
  - Objects added to a project are added to a default phase of CRISP-DM.
  - This means that you need to organize objects manually according to the data mining phase in which you used them.
  - It is wise to set the default folder to the phase in which you are currently working.

- **To select which phase to use as your default:**
  - In CRISP-DM view, right-click the folder for the phase to set as the default.
  - From the menu, choose Set as Default.
  - The default folder is displayed in bold type.

Clementine

Introduction

- The **Classes view** in the projects tool organizes your work in Clementine categorically by the types of objects created.

- Saved objects can be added to any of the following categories:

    - Streams

    - Nodes

    - Models

    - Tables, graphs, reports

    - Other (non-Clementine files, such as slide shows or white papers relevant to your data mining work)
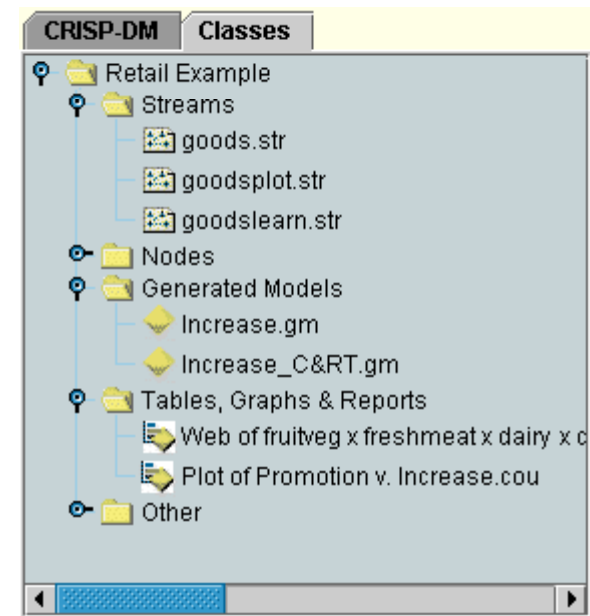
Building a Project

Building a Project

- A project is essentially a file containing references to all of the files that you associate with the project.

- This means that project items are saved both individually and as a reference in the project file (.cpj).

- Because of this referential structure, note the following:

  – Project items must first be saved individually before being added to a project.

  – Objects that are updated individually, such as streams, are also updated in the project file.

  – Manually moving or deleting objects (such as streams, nodes, and output objects) from the file system will make links in the project file invalid.

Creating a New Project

- **Creating a New Project**
  - You can either start building one, if none is open, or you can close an existing project and start from scratch.
- From the stream canvas menus, choose:
  - File > Project > New Project...

Clementine

Building a Project

- **Adding to a Project**
  - Once you have created or opened a project, you can add objects, such as data streams, nodes, and reports, using several methods.

Clementine

Adding Objects from the Managers

- **Adding Objects from the Managers**
  - Using the managers in the upper right corner of the Clementine window, you can add streams or output.

- **To Add Objects from the Managers**
  - Select an object, such as a table or a stream, from one of the managers tabs.
  - Right-click and choose Add to Project.
  - If the object has been previously saved, it will automatically be added to the appropriate objects folder (in Classes view) or to the default phase folder (in CRISP-DM view).
  - Alternatively, you can drag and drop objects from the managers to the project workspace.
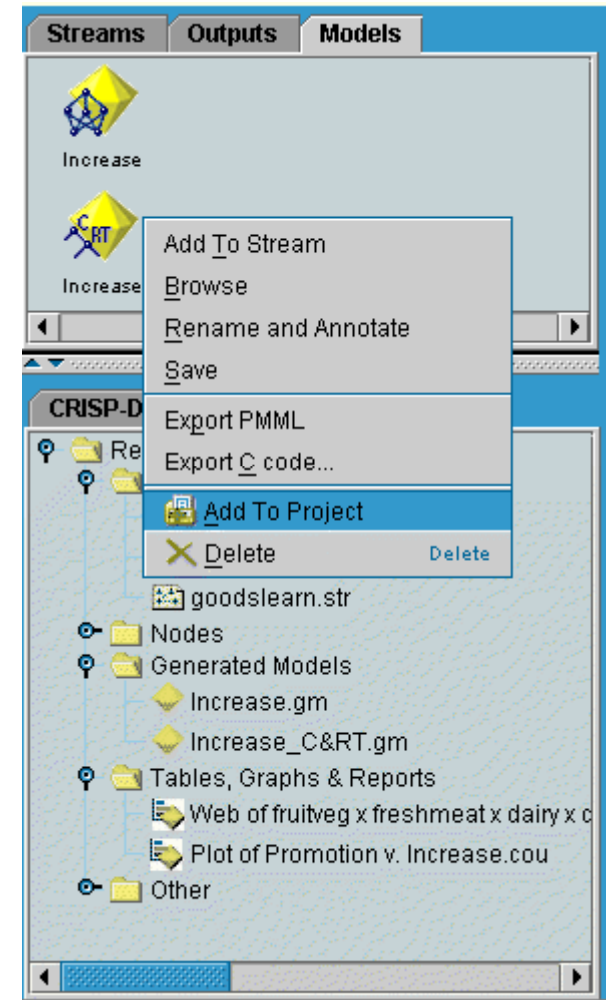
Adding Nodes from the Canvas

- **Adding Nodes from the Canvas**
  - You can add individual nodes from the stream canvas by using the Save dialog box.

- **To Add Nodes from the Canvas**
  - Select a node on the canvas.
  - Right-click and choose Save Node. Alternatively, from the menus choose: Edit > Node > Save Node...
  - In the Save dialog box, select Add file to project.
  - Create a name for the node and click Save.

- This saves the file and adds it to the project. Nodes are added to the Nodes folder in Classes view and to the default phase folder in CRISP-DM view.

Adding External Files

- **Adding External Files**

  - You can add a wide variety of non-Clementine objects to a project.

  - This is useful when you are managing the entire data mining process within Clementine.

  - For example, you can store links to data, notes, presentations, and graphics in a project.

  - In CRISP-DM view, external files can be added to the folder of your choice.

  - In Classes view, external files can be saved only to the Other folder.

Adding External Files

- **To add external files to a project:**
  - Drag files from the desktop to the project. or
  - Right-click the target folder in CRISP-DM or Classes view.
  - From the menu, choose Add to Folder.
  - Select a file in the dialog box and click Open.

- This will add a reference to the selected object inside Clementine projects.

Setting Project Properties

- **Setting Project Properties**
  - You can customize a project's contents and documentation by using the project properties dialog box.

- **To access project properties:**
  - ➤ Right-click an object or folder in the projects tool and choose Project Properties.
  - ➤ Click the Project tab to specify basic project information.

Building a Project

● Setting project properties

Folder Properties and Annotations

- **Folder Properties and Annotations**
  - Individual project folders (in both CRISP-DM and Classes view) can be annotated.
  - In CRISP-DM view, this can be an extremely effective way to document your organization's goals for each phase of data mining.
  - For example, using the annotation tool for the Business Understanding folder, you can include documentation such as "The business objective for this study is to reduce churn among high-value customers."
  - This text could then be automatically included in the project report by selecting the Include in report option.

- **To annotate a folder:**
  - Select a folder in the projects tool.
  - Right-click the folder and choose Folder Properties.
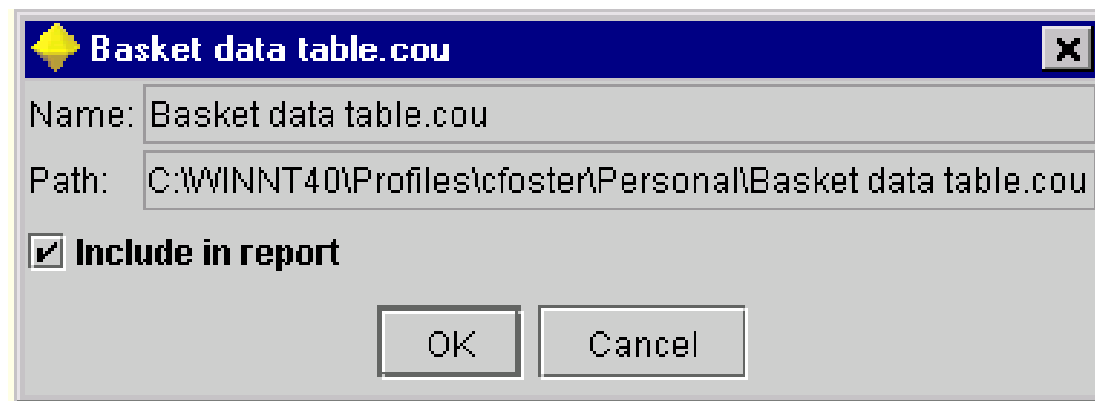
Clementine

Object Properties

- **Object Properties**
  - You can view object properties and choose whether to include individual objects in the project report.

- **To access object properties:**
  - Right-click an object in the project window.
  - From the menu, choose Object Properties.

Closing a Project

- **Closing a Project**
  - When you exit Clementine or open a new project, the existing project is closed, including all associated files.
  - Alternatively, you can choose to close the project file itself and leave all associated files open.

- **To close a project file:**
  - From the File menu, choose Close Project.
  - If you are prompted to close or leave open all files associated with the project, click Leave Open to close the project file (.cpj) itself but to leave open all associated files, such as streams, nodes, or graphs.

Clementine

References

References

- Integral Solutions Limited., **Clementine® 12.0 User's Guide**, 2007 (Chapter 6).

The end