HA-Lab 0: Preliminary Lab (self-learn)

Learning Objectives

This lab assignment aims to familiarize you with the installation of SAS EM, and creating a new project to begin data mining.

By the end of this lab, the students would be familiar with the process and terminology for:

1. Installation of SAS EM
2. Creation of a New Project
3. Creating a Library
4. Creating a Data Source
5. Creating a Diagram

The instructions of this lab are based on the SAS EM Tutorial available on SAS website. Click here to watch the video tutorial on YouTube.

Instructions

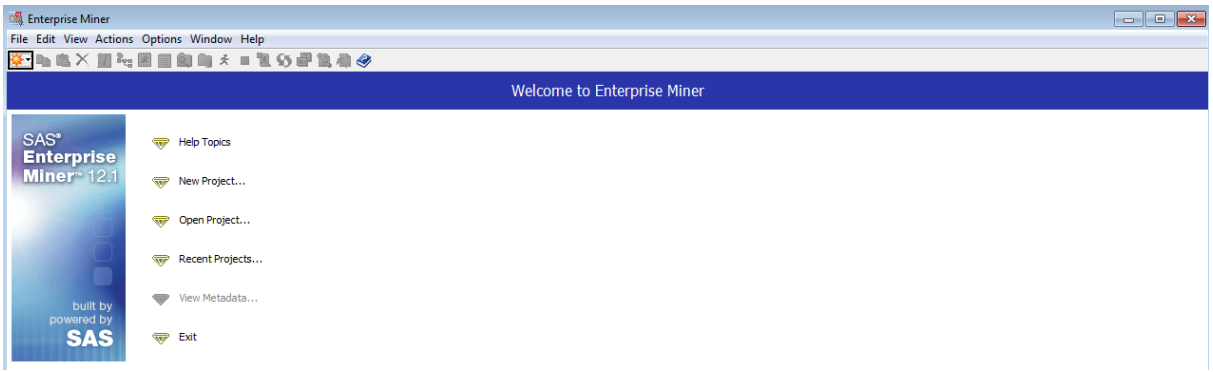
1. **Installation**

   SAS EM can be downloaded and installed from shared folder \\fs21\Applications\SAS_Installer\SAS_TM_EM

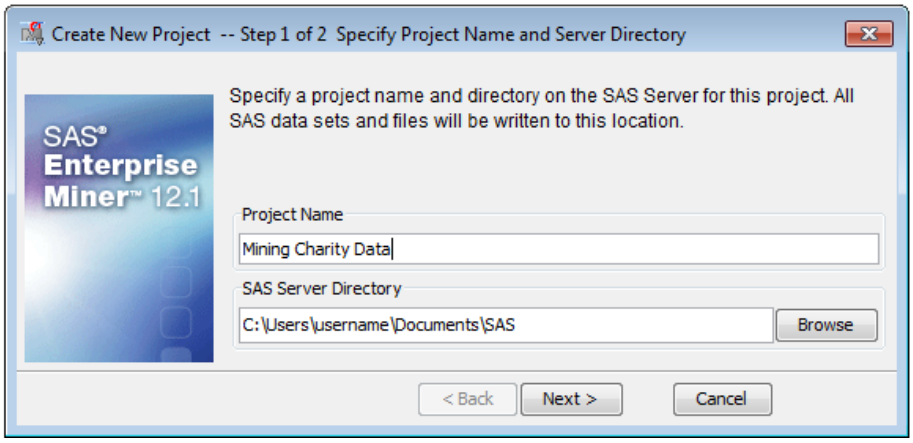
   Follow instructions found in "https://itservices.usc.edu/files/2014/09/installing-sas.pdf" "~~**SAS 9.3 TM EM Installation Guide.pdf**~~", and accordingly install the software. The entire installation procedure may take up to 2-3 hours.

2. **Create a new Project:** All work done for a data mining task is stored in a project in SAS EM

   a. After installation, launch "SAS Enterprise Miner Workstation 12.1". You should see the following window:



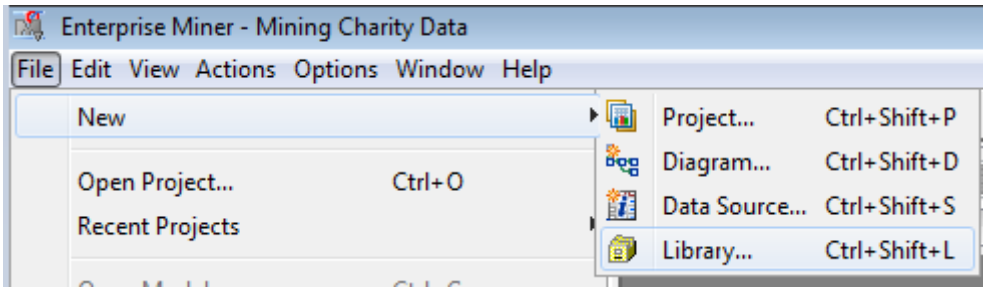
b. Click on **New Project,** give a project name, and select a local directory as the SAS Server Directory, and click **Next.**



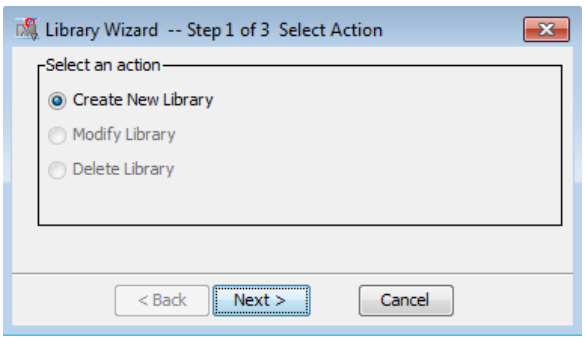
c. Click **Finish** in Step 2 of Creating a Project

3. **Creating a Library:** The library is used to store the sample data that you have downloaded from a source. These datasets are made accessible through this library.

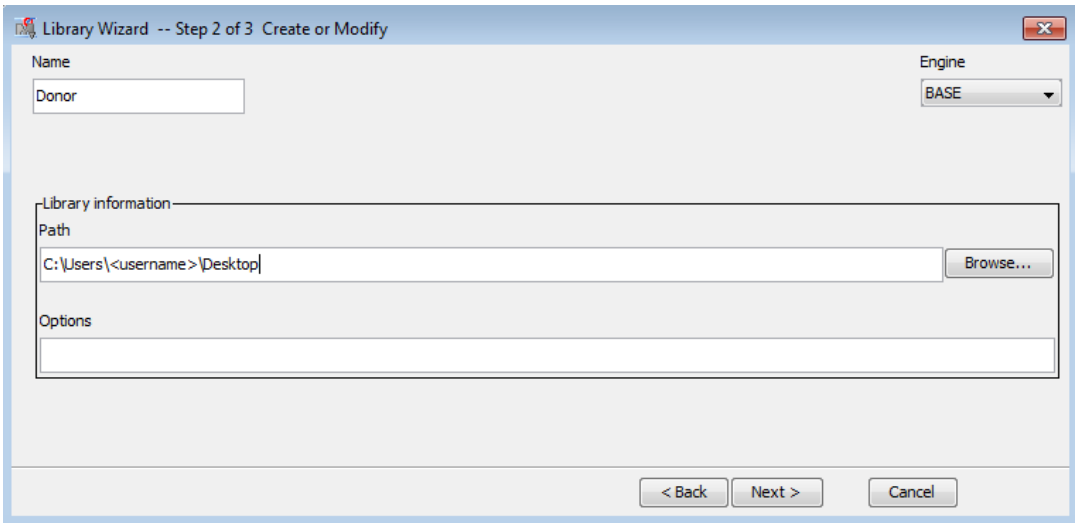
a. Click **File → New → Library.** The library wizard starts.



b. Select **Create New Library** and click **Next**



c. Select a name for the dataset, e.g. **Donor,** and provide the appropriate path where the downloaded data* was saved (e.g. you saved the dataset on your desktop)

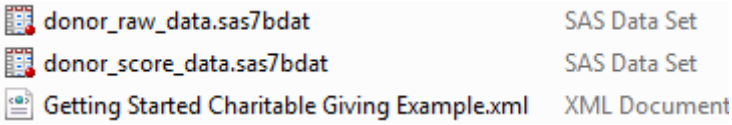


*You may download the sample data from
http://support.sas.com/documentation/onlinedoc/miner/#miner12x

SAS Enterprise Miner 12.1

- What's New in SAS Enterprise Miner 12.1 PDF
- Getting Started with SAS Enterprise Miner 12.1 PDF | HTML | Buy
- Example Data for Getting Started with SAS Enterprise Miner 12.1 ZIP
- SAS Enterprise Miner 12.1: Administration and Configuration PDF | HTML
- Developing Credit Scorecards Using Credit Scoring for SAS Enterprise Miner 12.1 PDF | HTML | Buy
- SAS Enterprise Miner 6, 7, and 12: C and Java Score Code Basics PDF
- SAS Enterprise Miner 12.1 Extension Nodes Developer's Guide PDF | HTML
- SAS Enterprise Miner 12.1: Reference Help, Second Edition (Secure Document)
- Help for SAS Enterprise Miner 12.1 is accessible within the product

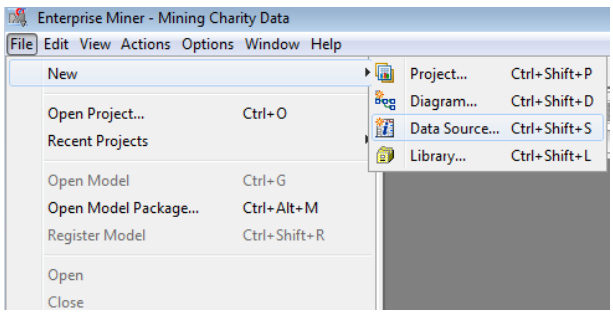
Download and unzip the "**Example Data for Getting Started with SAS Enterprise Miner".** When unzipped, you will see the following files:



d. Click **Finish.** The library has now been created.

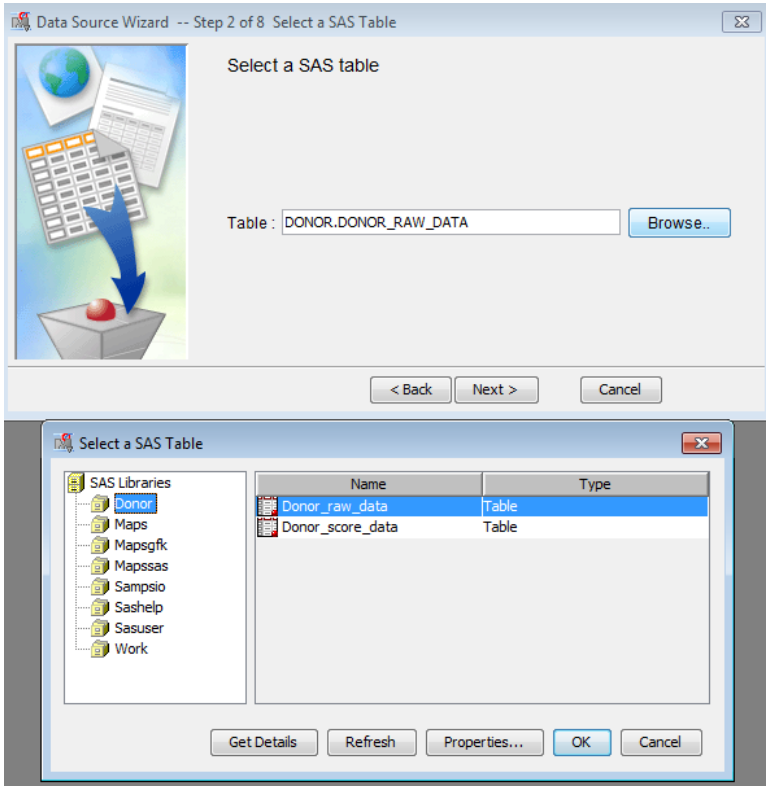
4. **Creating a Data Source:** In this step we enable the SAS EM to use the sample dataset.

a. Click **File → New → Data Source.** The Data Source wizard starts.



b. Click **Next** in step 1 (assuming SAS Table is automatically selected as source)

c. In step 2, Browse and select **DONOR_RAW_DATA** as the table, and click next



d. Click **Next,** Select the **Advanced** Option, and click **Next** again.

e. Now in step 5, we will decide what **role** is played by each feature. This is the same as deciding what is the target value, or the output (for example, something you want to predict), and what is the input value. For this assignment make the following changes to the roles:

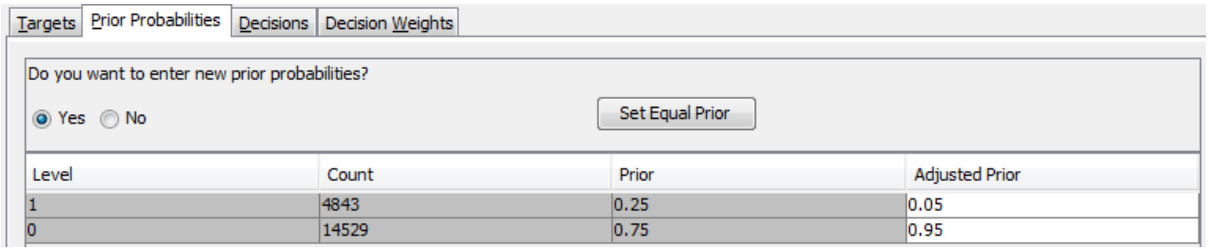
i. CLUSTER_CODE → REJECTED
ii. CONTROL_NUMBER → ID
iii. TARGET_B → TARGET
iv. TARGET_D → REJECTED
v. All other features → INPUT

An example is shown below

Name	Role
CARD_PROM_12	**Input**
CLUSTER_CODE	**Rejected**
CONTROL_NUMBER	**ID**

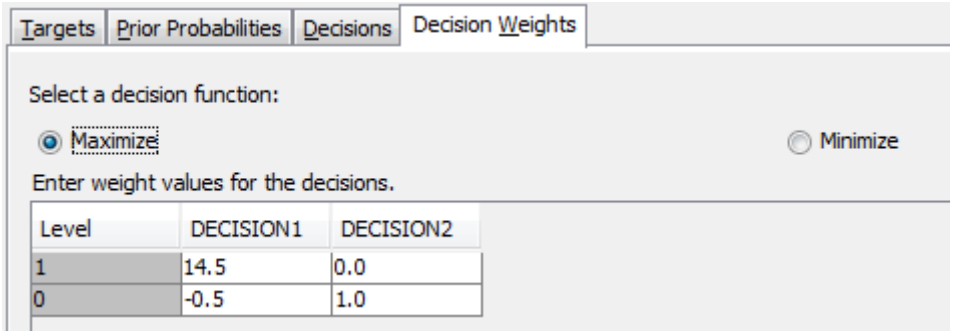
f. After assigning roles, and clicking **Next.** The Wizard asks whether the models built are based on decision making. Choose **Yes,** and click **Next.** The DONOR dataset is about decision making.

g. In step 7, on Decision Configuration, go to the **Prior Probabilities** tab, and select **Yes** where it is asked whether you want to enter new prior probabilities. Set the **Adjusted Prior** to 0.05 for Level 1, and 0.95 for Level 0.



The target value selected while setting the roles has 2 different values (e.g. binary classification). The prior reflects the fraction of data points belonging to each class. We adjust the prior based on some expert knowledge we have (for example in this case, an expert has told us that the prior distribution of the dataset is biased and the true distribution should be 0.05 and 0.95).

h. In the same step (step 70 now choose the **Decision Weights** tab.



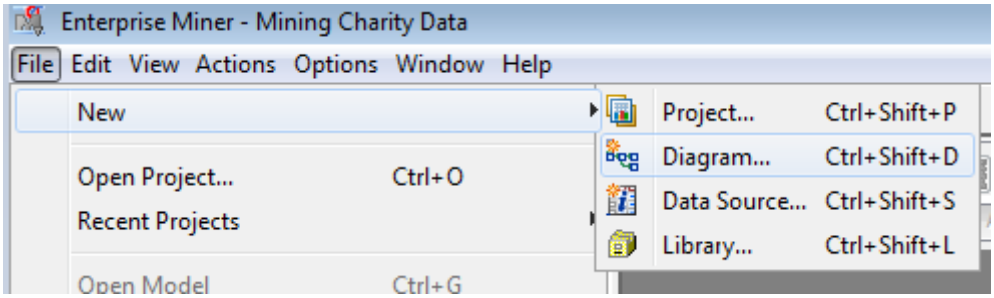
Modify the decision weights as shown in the image above, and ensure that the decision function is set to **Maximize.** These weights are assigning scores or rewards to the decisions taken given a particular scenario. For example, if Level 1 is the true scenario, we get a reward of 14.5 on taking Decision 1.

i. Click **Next,** Select **No** for sample data set, click **Next,** check that **Role** is set to **Raw.** Click **Next,** and then click **Finish.**

5. Create a Diagram and adding an input data node

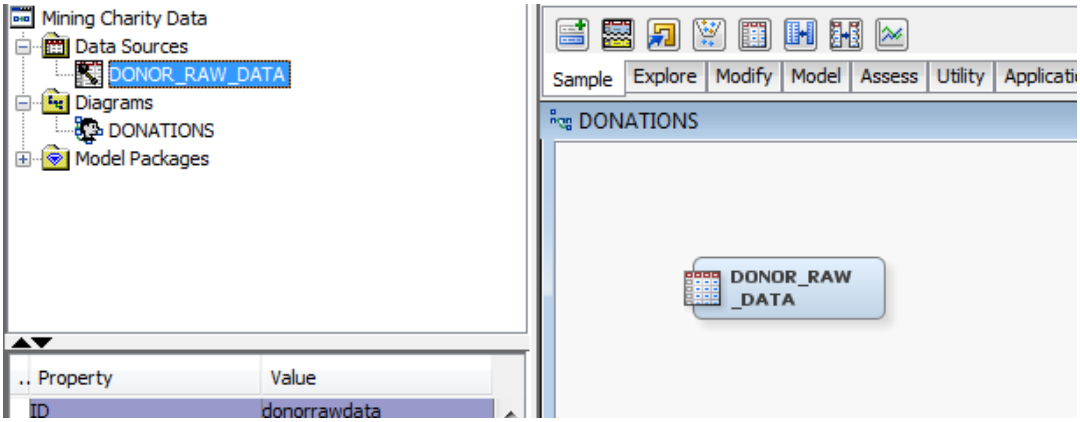
Now that we have set up our environment, we will take the first step towards a data mining process, by creating a process flow diagram.

a. Click **File → New → Diagram**



b. Enter "**DONATIONS"** as the diagram name, and click **OK.** An empty diagram opens in the **Diagram Workspace.**

c. Drag the DONOR_RAW_DATA from Data Sources onto your Diagram Workspace. The node gets created.



Exercise

There is no exercise or homework submission for this lab.