

Step1. Open the data/bank-data.csv Dataset

Click the “**Open file...**” button to open a data set and double click on the “**data**” directory.

Select the “bank-data.csv” file to load the bank dataset.

id	a unique identification number
age	age of customer in years (numeric)
sex	MALE / FEMALE
region	inner_city/rural/suburban/town
income	income of customer (numeric)
married	is the customer married (YES/NO)
children	number of children (numeric)
car	does the customer own a car (YES/NO)
save_acct	does the customer have a saving account (YES/NO)
current_acct	does the customer have a current account (YES/NO)
mortgage	does the customer have a mortgage (YES/NO)
pep	did the customer buy a PEP (Personal Equity Plan) after the last mailing (YES/NO)

Step2. Selecting or Filtering Attributes

In our sample data file, each record is uniquely identified by a customer id (the "id" attribute). We need to remove this attribute before the data mining step.

- We can do this by simply select the attribute and click on “Remove button”
- Using the Attribute filters in WEKA. In the "Filter" panel, click on the "Choose" button. This will show a popup window with a list available filters. Scroll down the list and select the "weka.filters.unsupervised.attribute.Remove" and click apply.
- Save the file with a name "bank-data-R1.arff"

Step3: Discretization

Some techniques, such as association rule mining, can only be performed on categorical data. This requires performing discretization on numeric or continuous attributes. There are 3 such attributes in this data set: "age", "income", and "children". In the case of the "children" attribute the range of possible values are only 0, 1, 2, and 3. In this case, we have opted for keeping all of these values in the data. This means we can simply discretize by removing the keyword "numeric" as the type for the "children" attribute in the ARFF file, and replacing it with the set of discrete values. We do this directly in our text editor. In this case, we have saved the resulting relation in a separate file "bank-data2.arff".

Load the bank-data2.arff dataset into weka.

If we select the "children" attribute in this new data set, we see that it is now a categorical attribute with four possible discrete values.

once again we choose the Filter dialog box, but this time, we will select Discretize from the list. Next, to change the defaults for this filter, click on the box immediately to the right of the "Choose" button. This will open the Discretize Filter dialog box. We enter the index for the attributes to be discretized. In this case we enter 1 corresponding to attribute "age". We also enter 3 as the number of bins (note that it is possible to discretize more than one attribute at the same time by using a list of attribute indices).

Save the file as bank-data3.arff and check the dataset in text editor.

Next, we apply the same process to discretize the "income" attribute into 3 bins. Again, Weka automatically performs the binning and replaces the values in the "income" column with the appropriate automatically generated labels. We save the new file into "bank-data3.arff", replacing the older version. Editing by text editor made bank-data-final.arff.

Step4: Missing Values

1. Open file "bank-data.arff"
2. Check if there is any missing values in any attribute.
3. Edit data to make some missing values.
4. Delete some data in "region"(Nominal) and "children"(Numeric) attributes. Click on "OK" button when finish.
5. Make note of Label that has Max Count in "region" and Mean of "children" attributes.
6. Choose "ReplaceMissingValues" filter
(weka.filters.unsupervised.attribute.ReplaceMissingValues). Then, click on Apply button.

7. Look into the data. How did those missing values get replaced ?
8. Edit “bank-data.arff” with text editor. Make some data missing by replacing them with ‘?’. (Try with nominal and numeric attributes). Save to “bank-data-missing.arff”.
9. Load “bank-data-missing.arff” into WEKA, observe the data and attribute information.
10. Replace missing values by the same procedure you had done before.