**Using WEKA for creating models and obtaining predictions**

Michael Towsey

18th April 2018

This information was valid for Weka version 3.8.1.

# Creating Models

This is relatively straight forward, and you can check some of the following websites.

https://machinelearningmastery.com/save-machine-learning-model-make-predictions-weka/

Note: Weka can train a model even if the dataset contains instances that do not have a valid label. Such instances must be labelled with a “?”.

# Saving models

This is relatively straight forward. These websites are helpful

https://weka.wikispaces.com/Saving+and+loading+models

# Loading Models

This is relatively straight forward. These websites are helpful

https://weka.wikispaces.com/Saving+and+loading+models

IMPORTANT NOTE: Before you can access the Classify Tab where you make, save and load models, you *must* load a data set in the Preprocess panel. If your intention is to use the model to make predictions, you can load any old data set or dummy data file. The dummy data will not be used. It is required only to gain access to the Classify panel.

1: Right click in the Classify/Result List panel in order to load a pre-saved model.

2: Once you have selected a model, its name will appear in the “Results list” panel. Right click on the model’s name and click “Re-evaluate model on current test set”. If this option is not accessible, you need first to load a Supplied Test Set.

# Making Predictions

https://machinelearningmastery.com/save-machine-learning-model-make-predictions-weka/

When making predictions with a model using a supplied test file:

1: click the “Supplied Test Set” button in the “Classify/Test Options” panel.

2: Click the “Set…” button. This will open a “Test Instances” dialogue box.

3: Click the “Open file…” button to open a test file. When you have selected the file, the “test Instances/Class” button should show the correct name of the Label category.

4: Ignore the name of the target class that appears in the button below the Classify/Test Options panel. This refers to the data file that was loaded through the Preprocess tab (and is being ignored).

5: IMPORTANT NOTE: The format of the test file MUST be same as that of the training data file. So, if the training file contained N columns the “Supplied Test Set” file must also contain N columns, having the same column header names.

It appears that it is not possible to edit the file contents of the “Supplied Test Set” file. This has a couple of consequences. In particular:

1. It appears necessary to remove any unused attribute columns from the final “Training Data” file, because you cannot have them in the “Supplied Test Set” file. For example, if your “Training Data” file contained a column displaying the instance ID number, remove it because it will complicate the later prediction process.
2. The “Supplied Test Set” file can contain question-marks, “?”, where the instance label is unknown.

6: Before clicking the “Start” button, click the “More Options…” button to set a bunch of prediction options. Turn off (unclick) the following check-boxes because they are not necessary for prediction:

a: Output model

b: Output per class stats

c: output entropy evaluation.

d: error plot point size

e: etc.

7: In order to save your predictions to a csv file, click the “Choose” button beside “Output Predictions” and then click the Output/Prediction/CSV option.

8. Left click over the word “CSV” and left click in the “outputFile” text box. Enter the name of the file to which the output will be saved. I have always accepted the default options for the rest of the dialogue.

9: Return to the Classify panel and click the “Start” button.

10: The predictions should be written to the CSV file you nominated. Check the above internet sources for the formatting of the predictions output file.