

## Weka Online (last updated: 08-08-2008)

The test option is 10-fold stratified cross-validation (exactly as implemented by default in Weka, using default seed '1' of Weka 3-5-7). We as do many other machine-learning practitioners (see Publications) consider this reliable enough as well as economical. We here go through the stages:

**1. Registration and login.** Follow the steps. You will receive account validation and confirmation messages. After that you are directed to dataset submission page. (Note that the only way to access submission and results pages is to login first)

**2. Submission page.** (There is now a link to getting results at the top of the page)

- **Dataset** should be valid ARFF file. Dataset's first attribute should be an instance ID. This will not only enable the tracking down of instances for errors but also the we'll know to filter out this attribute from being used as a feature in classification so as not to get skewed results! Tip: Use Weka's *AddID* filter in Preprocess tab to add that ID.
- **Classifiers** (textarea) is for adding classifiers in Weka syntax. There should be a line change between classifiers and no empty line as last line. Tip: Right-click (Copy configuration to clipboard) on classifier field in Classify tab of Explorer, and paste that here. We also provide a list of strong classifiers.
- **Project and usernames** should not contain the reserved character '/' nor whitespaces ' '. Results are stored under username -> project (and you can have multiple accounts of both). Note that you cannot query for results if you don't give a project name! You should use the same username as you registered with. Email address is here only for sending results via email to you (so you don't have to stand by online for the project to finish!). See Get the results below how to query without that.

After submit, program asks for username and password to run Weka Online script. These should have been provided to you in account validation message.

**4. Get the results.** You can query the results of projects here. We have made available these:

- Classifier accuracies at datasets
- Errors at instances by classifiers
- Dataset factor breakdown (number of classes and instances in each of them)
- ROC, TP and FP rates per class

## Interpreting the results query outputs

Here's a sample output from **Accuracies** form:

### Results

Thank you . You submitted /queried for a project called **24sf-ClassesSfs** in **results summary mode**. Here are the results:

Project	Class	Classifier	Accuracy
24sf-ClassesSfs		weka.classifiers.meta.END -S 1 -I 10 -W weka.classifiers.meta.nestedDichotomies.ClassBalancedND -- -S 1 -W weka.classifiers.meta.Bagging -- -P 100 -S 1 -I 10 -W weka.classifiers.trees.J48 -- -M 2 -C 0.25	92.670
24sf-ClassesSfs		weka.classifiers.meta.END -S 1 -I 10 -W weka.classifiers.meta.nestedDichotomies.ClassBalancedND -- -S 1 -W weka.classifiers.meta.Bagging -- -P 100 -S 1 -I 10 -W weka.classifiers.rules.PART -- -M 2 -C 0.25 -Q 1	92.670
24sf-ClassesSfs		weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -SBAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5	91.099
24sf-ClassesSfs		weka.classifiers.functions.LibSVM -S 0 -K 0 -D 1 -G 1.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.0010 -P 0.1	89.537
24sf-ClassesSfs		weka.classifiers.functions.LibSVM -S 0 -K 1 -D 2 -G 1.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.0010 -P 0.1	88.679
24sf-ClassesSfs		weka.classifiers.bayes.NaiveBayes	87.784

Classification task in this case is *the recognition of protein class* (6 classes) for 24 protein superfamilies (i.e. actual proteins with certain function).

- First column is obviously the given project name.
- Second is the name of the classification (sub)task, which in Weka Online (which does not allow specification of a class structure file) is always null.
- Classifier is the name of the Weka runnable (we will have a utility where you can specify a shorter name).
- Accuracy is the classification accuracy (number of instances correct / total number of instances) of the classifier on the subtask. Classifiers are ranked in descending order of accuracy.
- So in this case Weka Online ranks END classifier the highest.

## Support and troubleshooting

If it does not work, contact us at [developers@ceodelegates.com](mailto:developers@ceodelegates.com).