

Université de Carthage	Module : Data Mining
INSAT	Sections : GL
Département Mathématiques et Informatique	Niveau : 4ème année
Année Universitaire : 2023 - 2024	Enseignante : Sana Hamdi

TP N°2 : Introduction à Weka Experimenter

Objectif :

Ce TP a pour but de vous familiariser avec le Weka Experiment Environment. En effet, cet environnement va vous permettre de créer, lancer, modifier et analyser des expériences de manière plus souple que l'utilisation des classifieurs individuellement. Par exemple, vous pouvez créer une expérience qui lance plusieurs classifieurs sur des séries de jeux de données et analyser les résultats pour déterminer si l'un des classifieur est statistiquement meilleur que les autres.

I- Définir et sauvegarder une expérience

Quand l'Experimenter est démarré, la fenêtre Setup est affichée. La première étape consiste à choisir un jeu de données de travail.

1. Cliquez sur New (en haut à droite de la fenêtre Setup) pour initialiser une expérience. Ceci permet de fixer les paramètres par défaut pour l'expérience.
2. Afin de ne pas refaire le même travail plusieurs fois et de pouvoir exploiter les résultats de l'expérience, on peut sauvegarder les résultats de l'expérience dans un fichier. Pour identifier un fichier vers lequel les résultats seront envoyés, dans la fenêtre « Results Destination » (en haut de la fenêtre Setup), se placer sur « browse » de l'output Filename. Cliquez sur ce paramètre pour afficher une fenêtre de sélection de fichiers. Tapez le nom du fichier de sortie (output), par exemple Experiment.csv ou le sélectionner via open, cliquez sur Select, et cliquez sur close (X). Vérifiez que tous les fichiers de sortie sont situés dans votre répertoire (et pas dans le répertoire weka).

3. Pour définir le jeu de données (dataset) à traiter par un classifieur, sélectionnez d'abord *Use relative paths* dans la fenêtre Datasets (en bas à gauche de la fenêtre Setup) et cliquez sur Add new pour ouvrir une fenêtre de dialogue. Sélectionnez le dataset Iris. Le nom du dataset est maintenant affiché dans la fenêtre Datasets de la fenêtre de Setup.

II- Lancer une expérience

Sans autre indication, c'est l'expérience par défaut est exécutée sur le dataset que vous avez choisi, soit, normalement le CrossValidation avec un paramètre à 10 pour l'apprentissage aléatoire et des tests effectués sur le dataset choisi (Iris).

Choisissez le classifieur ZeroR.

Pour lancer l'expérience que vous venez de configurer, cliquez sur l'onglet Run en haut de la fenêtre Experiment Environment et cliquez sur Start. L'expérience est exécutée.

```
Started
Finished
There were 0 errors
```

La ligne 1 contient les identifiants de colonnes (du résultat de l'expérience). Chaque ligne définit un jeu d'apprentissage et un test. La ligne 2 de la feuille de données indique que pour le premier run de l'expérience, le dataset Iris a été utilisé avec le classifieur ZeroR et que W instances ont été testées par le classifieur : X instances ont été classées correctement, Y instances ont été mal classées, et Z instances n'ont pas pu être classées.

Laquelle (lesquelles) de ces colonnes pourrai(en)t être utile pour analyser l'efficacité d'un classifieur ?

III- Comparer les classifieurs

Pour comparer plusieurs classifieurs, nous devons les ajouter dans le panneau des algorithmes.

Dans le panneau Algorithms (en bas à droite de la fenêtre Setup), cliquez sur Add New et ajoutez les classifieurs OneR, ZeroR et J48. (Pour enlever un classifieur, sélectionnez-le en cliquant dessus et cliquez Delete selected).

Lancez l'expérience et regardez dans la fenêtre Analyse, cliquez sur Experiment puis Perform Test.

Analysez les résultats..

IV- Ajouter des jeux de données

Le(s) classifieur(s) peuvent être lancés sur un nombre quelconque de datasets à la fois. Les datasets sont ajoutés en cliquant sur Add new dans le panneau Datasets (en bas à gauche de la fenêtre Setup). Les datasets sont effacés de l'expérience en sélectionnant le dataset voulu et en cliquant sur Delete selected.

Ajoutez 5 différents jeux de données et Analysez les résultats..

V- Compte Rendu :

1. Soit le dataset banque.csv. appliquez 3 prétraitements différents.
2. Créez une expérience qui lance plusieurs classifieurs (J48, RandomForest, IBK k=1, IBK k=3, IBK k=5, IBK k=10) sur la série de jeux de données.
3. Analysez les résultats.