Le 09/12/2022

Réalisé par :

Safa Laabidi

Mohamed Rayen Samet

Skander Menzli

Farouk Drira

GL4

**Compte rendu du TP n°2**

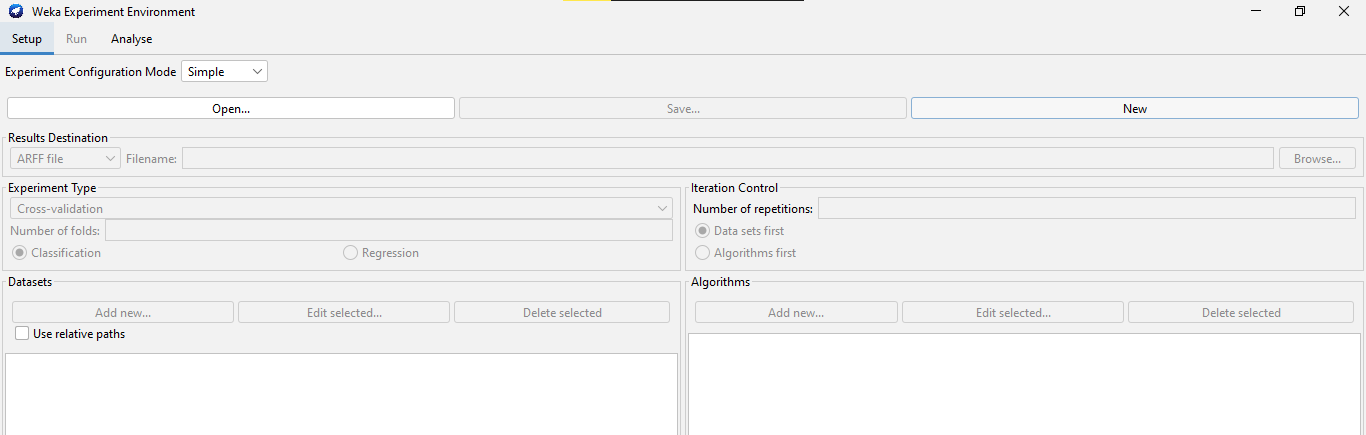
**DM & ML**

****

Se Familiariser avec Weka Experiment Environment

1. Création de l’expérience :

On va créer une expérience qui lance plusieurs classifieurs (J48, RandomForest, NaiveBayes, IBK k=1, IBK k=3, IBK k=5, IBK k=10) sur la série de jeux de données suivante : IRIS, Glass et Weather.



Une image contenant texte

Description générée automatiquement

On clique sur l’onglet « Run » pour obtenir un fichier ARFF ou CSV.

Une image contenant texte

Description générée automatiquement

Une image contenant table

Description générée automatiquement

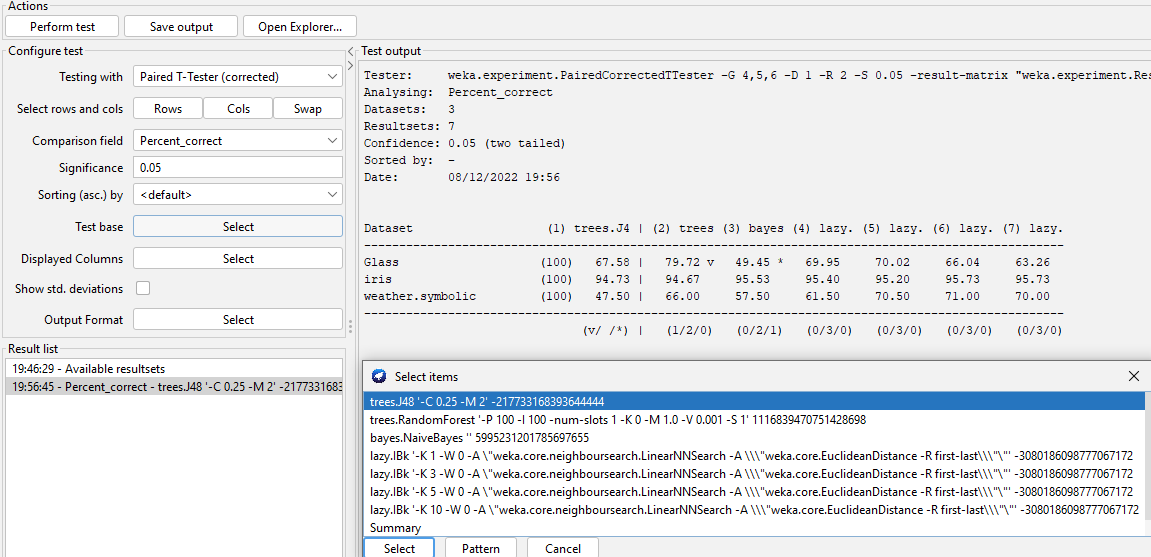
1. Analyse :

Ensuite on clique sur l’onglet « Analyze », puis le bouton « Experiment ».

Une image contenant texte

Description générée automatiquement

En cliquant sur « Perform Test », on obtient une comparaison par rapport au pourcentage d'instances **correctement classées** (par défaut). Tous les algorithmes sont comparés par rapport aux performances du classifieur J48 avec un **intervalle de confiance** égal à 0.05 (par défaut).



Une image contenant texte, table

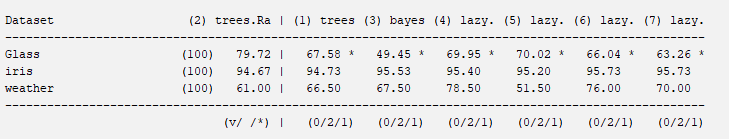
Description générée automatiquement

Les classifieurs utilisés dans l'expérience sont indiqués dans les colonnes et les datasets dans les lignes. Le pourcentage des instances correctement classifiées pour chacun des 7 algorithmes est indiqué dans chaque jeu de données.

L'annotation v ou \* indique qu'un résultat est statistiquement meilleur (v) ou pire (\*) que le classifieur de référence (dans ce cas, J48).

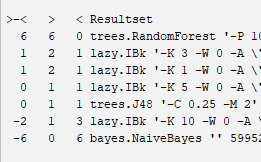
**Autre exemple :**

Tous les algorithmes sont pires que « Random Forest » sur le dataset Glass.



**Ranking Test :**

En changeant la base de test à « Ranking » on obtient un classement de tous les classifieurs en fonction du nombre total de gains (>) et de pertes (<) par rapport aux autres classifieurs. La première colonne (> − <) est la différence entre le nombre de victoires et le nombre de défaites. Cette différence est utilisée pour générer le classement.



On remarque que « Random Forest » a performé mieux que tous les autres classifieurs et a obtenu le meilleur score (= 6 ), et Naïve-Bayes a obtenu le score le plus bas (= -6 ).