Université Sultan Moulay Slimane
Ecole National des Sciences Appliquées
Khouribga

Module : Data Mining & Texte Mining

1ère Année Master Big Data & Aide à la Décision
Semestre 2 / Année universitaire 2018/2019
Feuille de Travaux Pratiques N° 1

OBJECTIF DE L'ACTIVITE PRATIQUE :

*Dans ce TP vous allez apprendre le logiciel open source **KNIME** pour le processus du Data Mining (Accès aux données, Transformation de données, Analyses prédictives et Visualisation).*

Prise en main du logiciel KNIME

1. Lisez attentivement la page Workbench User Guide.
2. Reconnaissez la terminologie de base de KNIME : **nœud**, **espace de travail**, **flux de travail**, **état d'un nœud**, **connexion** et **configuration de nœuds**.
3. Installez le logiciel **KNIME** sur vos ordinateurs, lancez-le, puis spécifiez un dossier de travail (**workspace**) où vous stockerez vos projets comme indiqué dans la figure 1 :

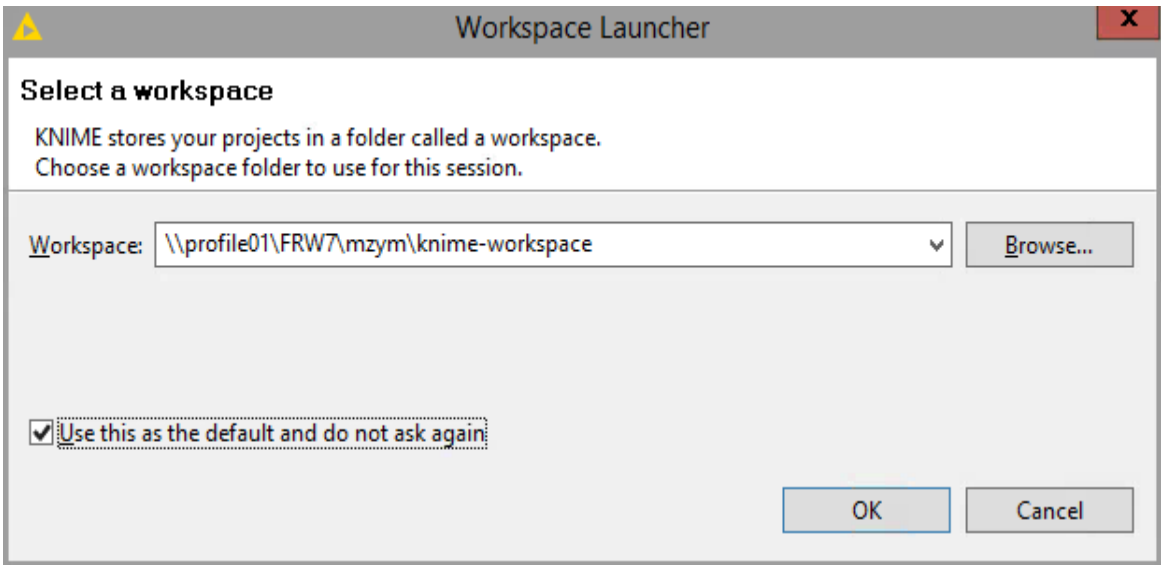

Fig. 1. Spécification du chemin d'accès de l'espace de travail (workspace).

Création et exécution d'un flux de travail étape par étape

L'objectif de cette activité est la création et l'exécution, étape par étape, d'un flux de travail (**workflow**) simple. Ce dernier lisera les données à partir d'un fichier Excel, effectuera un clustersing de ces données et les affichera sous diverses formes.

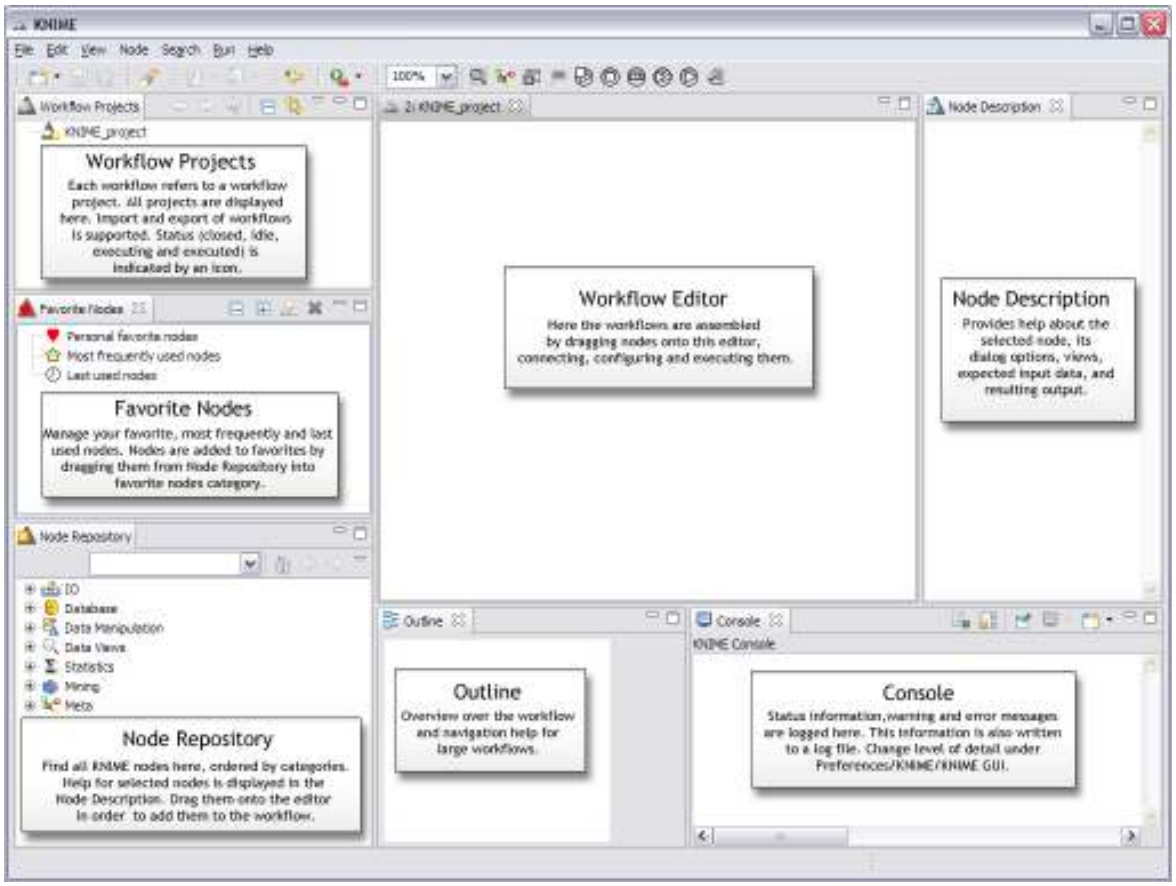

Fig. 2. L'écran du logiciel KNIME.

1. Lancez **KNIME** (assurez-vous qu'il démarre avec un workflow vide).
2. Ajoutez un nœud "**Read**" au workflow comme suit. Dans l'explorateur des nœuds, cliquez sur la catégorie "**IO**", puis sur la sous-catégorie "**Excel Reader (XLS)**" comme illustré dans la figure 3 (écran à gauche). Ensuite, faites un "glisser-déplasser" de l'icône "**Excel Reader (XLS)**" dans la fenêtre de l'éditeur de Workflows.

Fig. 3. Le nœud « Excel Reader (XLS) ».

3. En procédant de la même manière que précédemment, ajoutez les nœuds "**k-Means**", "**Color Manager** ", "**Interactive Table** " et "**Scatter Plot**" au même workflow. Vous devez obtenir le même suivant résultat comme indiqué par la figure 4.

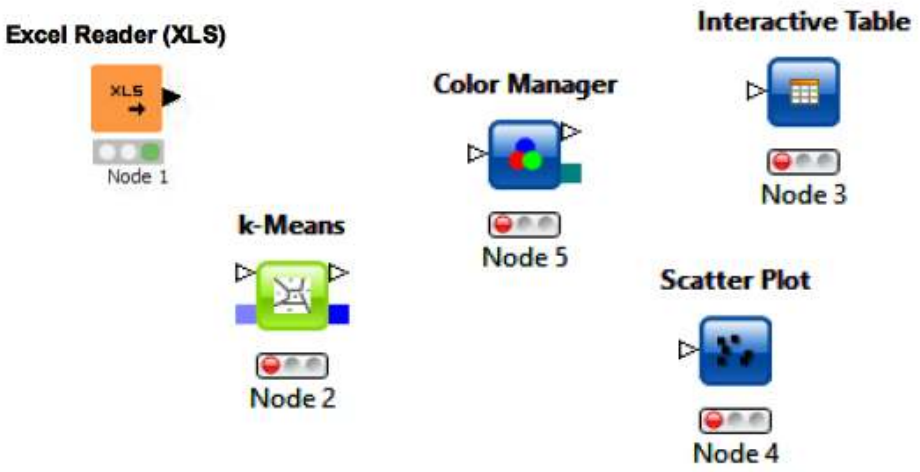

Fig. 4. Un simple workflow avec 5 nœuds non connectés.

4. Connectez les nœuds, en cliquant sur le port de sortie d'un nœud et en le-tirant jusqu'à le port d'entrée du nœud concerné, afin d'obtenir le résultat de la figure 5. Notez que tous les nœuds ne montrent pas l'état vert puisqu'ils ne sont pas configurés et exécutés.

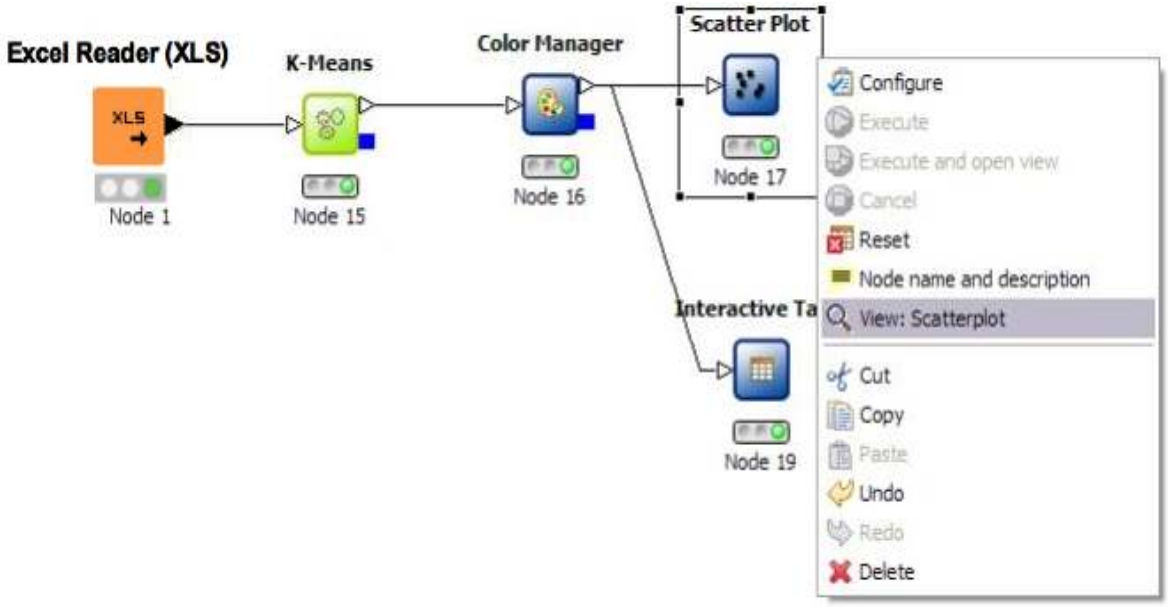

Fig. 5. Le workflow avec les nœuds connectés et exécutés.

5. Faites un double-clique sur le nœud "**Excel Reader (XLS)**" et sélectionnez la commande "**Configure**" à partir du menu contextuel. Dans la boîte de dialogue qui vient de s'ouvrir, sélectionnez le fichier Excel "Iris.xls" comme source de données.

6. Pressez le bouton de commande "**OK**" pour fermer la boîte de dialogue de configuration du nœud "**Excel Reader**". Une fois le nœud est correctement configuré, son état passé au jaune (signifiant que le nœud est prêt pour exécution). Ensuite, le nœud "**K-Means**" va immédiatement passer à la couleur jaune, puisque ses paramètres de configuration par défaut seront appliqués automatiquement. Pour être sûr que ces paramètres correspondent bien à vos besoins, ouvrez la boîte de configuration de ce nœud et inspectez les paramètres par défaut.
7. Pour configurer le nœud "**Color Manager**", vous devez exécuter le nœud "**K-Means**". Après son exécution, toutes les valeurs nominales et les intervalles de tous les attributs sont connues : cette information sera propagée aux nœuds successeurs. Une fois le noeud "**K-Means**" est exécuté, ouvrez la boîte de configuration du nœud "**Color Manger**" (voir le figure 6).

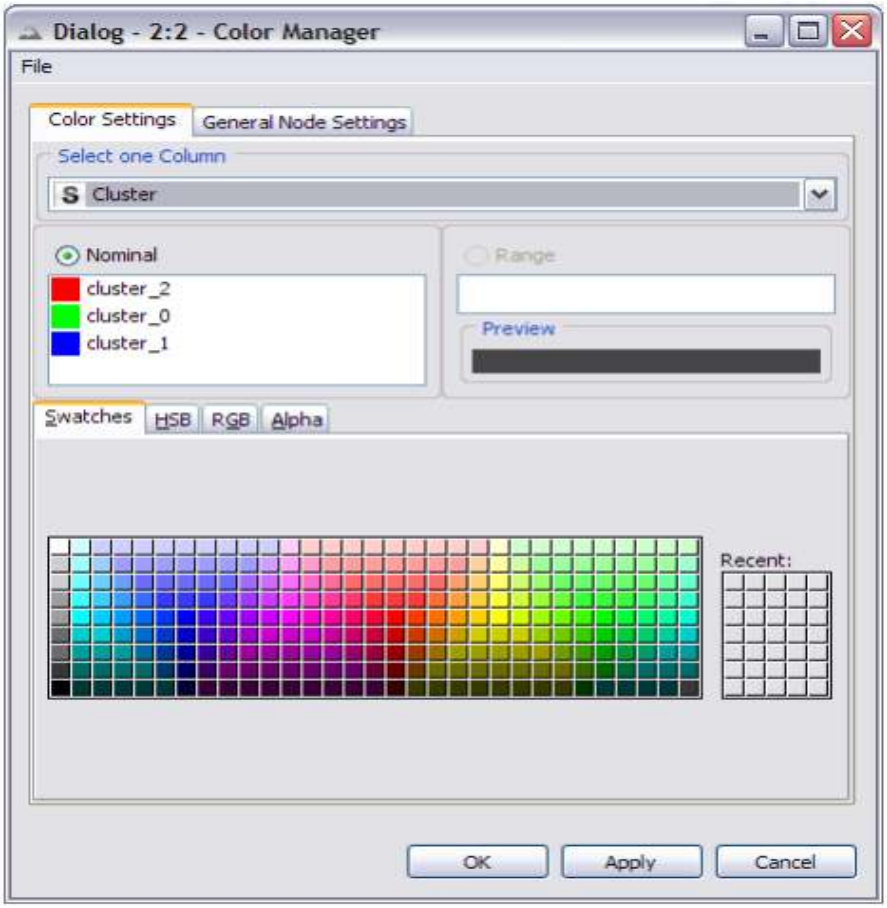

Fig. 6. La boîte de dialogue de configuration du nœud "Color Manager".

8. Exécutez le nœud "**Scatter Plot**", **KNIME** exécutera alors tous les nœuds prédécesseurs. Dans un workflow large et plus complexe, vous devez parfois sélectionner plusieurs nœuds à la fois et les exécuter tous. Dans ce cas, le gestionnaire du workflow exécutera les nœuds nécessaires, si possible en parallèle.

9. Ouvrez Les aperçus des nœuds “**K-Means**”, “**Interactive Table**” et “**Scatter Plot**” en utilisant leurs menus contextuels.

10. Sélectionnez quelques points dans le “**scatter plot**” et choisissez la commande “**Hilite Selected**” à partir du menu “**Hilite**”. Les points mis en relief sont marqués d’une bordure orange. Vous pouvez également voir les points mis en relief dans l’aperçu “**table view**”. La propagation de l’état “**hilite**” fonctionnera pour tous aperçus dans toutes les branches du flux affichant les mêmes données.

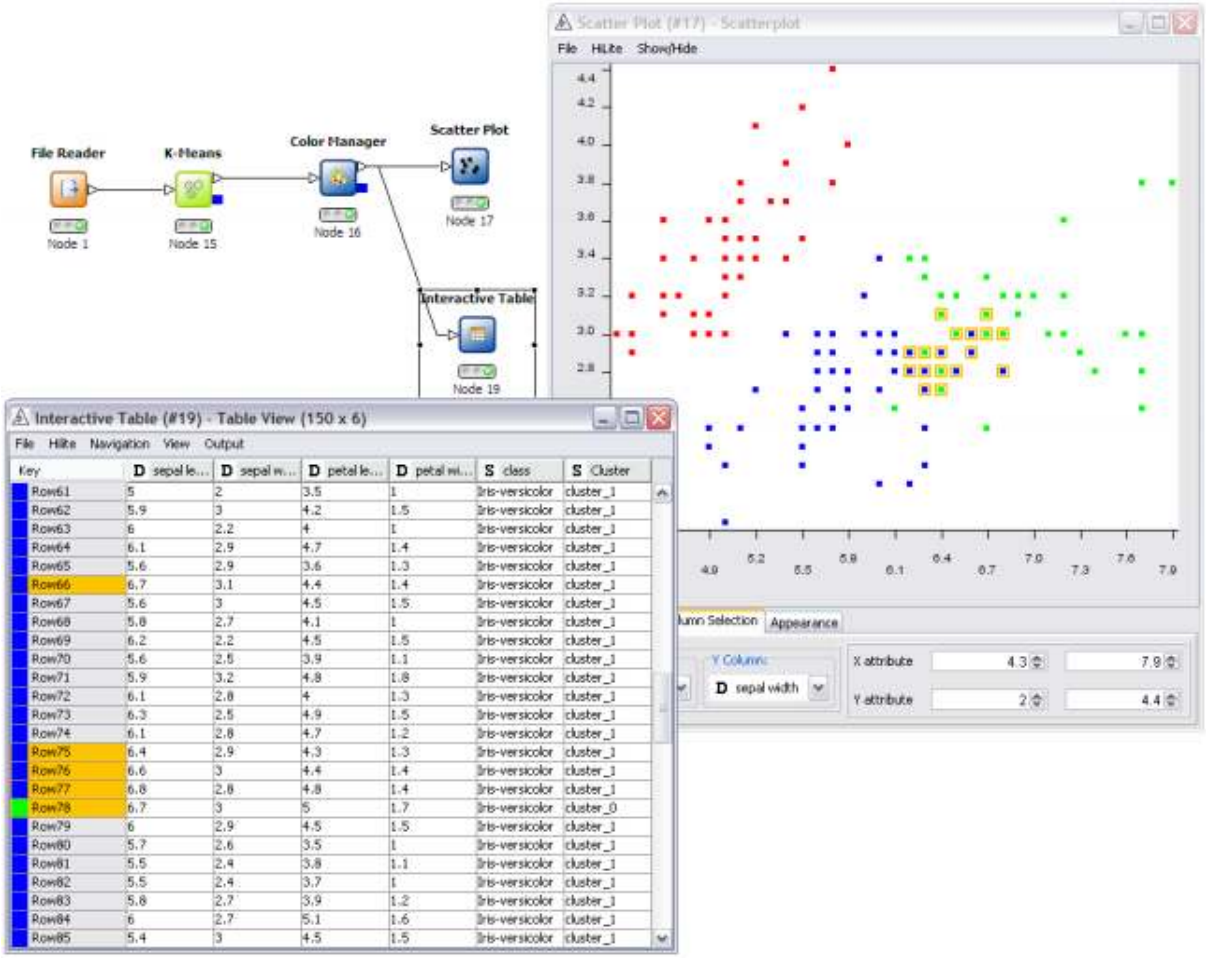

Fig. 7. Mis en relief des données sélectionnées.