



## Compte Rendu d'Articles

### HMIN326

### 2020 - 2021

Le sujet est individuel. Choisissez un article parmi les 3 proposés sur Moodle puis répondez aux questions associées.

Consignes :

- Vous devez déposer sur Moodle un fichier au format pdf identifié par le numéro de l'article retenu suivi de votre numéro d'étudiant comme l'illustre l'exemple suivant : article1\_20192342 où article1 correspond à l'article 1 de la liste et 20192342 à votre numéro d'étudiant.
- Le fichier doit contenir votre prénom, nom et numéro d'étudiant ainsi que de manière explicite les réponses aux différentes questions.

#### Article 1 :

Jianyong Wand and Jiawei Han. "BIDE: Efficient Mining of Frequent Closed Sequences". In Proceedings of the 20th International Conference on Data Engineering (ICDE 2004).

1. Résumer l'article en précisant les principales contributions (une demi-page max).
2. A l'aide de la base de données exemple de la Table 1, illustrer les différentes étapes de l'algorithme Bide (il s'agit de faire la trace complète de l'algorithme).
3. Quelle est la différence entre des séquences maximales et des séquences fermées (*closed sequences*) ?
4. A votre avis est-ce que le nombre de séquences maximales est supérieur ou inférieur au nombre de séquences closes ? 'itemsets clos' ? Justifier la réponse.
5. Dans le cadre de votre projet de classification vous avez utilisé différents traitements. Préciser comment l'utilisation de motifs séquentiels pourrait aider à améliorer les classifieurs. Indications : les items peuvent être considérés comme des mots pré-traités et la classe est connue.

#### Article 2 :

Dino Ienco, Ruggero G. Pensa and Rosa Meo: "Context-based Distance Learning for Categorical Data Clustering". In Proceedings of the 8th Intelligent Data Analysis conference (IDA 2009) [PDF](#)

1. Résumer l'article en précisant les principales contributions (une demi-page max).
2. Expliquer comment le contexte est sélectionné dans la méthodologie DILCA.
3. Détailler le protocole expérimental utilisé et expliquer quels sont les aspects de la méthodologie proposée que les expériences visent à évaluer.
4. Quel est l'impact du paramètre sigma sur les performances de l'algorithme ?





5. Dans l'analyse des performances en temps, quels sont les composantes de l'approche DILCA qui sont évalués et comment vous expliquez les différences en terme de coût computationnel des différentes étapes?

**Article 3 :** Waleed Ragheb, Jérôme Azé, Sandra Bringay, Maximilien Servajean.

“Pourquoi dois-je croire ta prédiction ?” In *30es Journées Francophones d'Ingénierie des Connaissances (IC)* (pp. 112-124). [PDF](#)

1. Résumer l'article en précisant les principales contributions (une demi-page max).
2. A la base de l'article, expliquer le concept d'“explicabilité” des systèmes d'intelligence artificiel et en particulier des modèles d'apprentissages automatique.
3. Comment les auteurs se positionnent par rapport à l'état de l'art dans le domaine étudié ? En quoi l'approche proposée est différente (originale) par rapport à l'existant ?
4. Quelle est la différence entre les deux types d'expériences qui ont été fait ? Que voudrait démontrer chacun des deux tests ?
5. Donner trois faiblesses de l'article et trois points forts.

