

Rendu article

HMIN232 - 2019-2020

Choisissez un article parmi les 4 proposés sur Moodle puis répondez aux questions associées.

ARTICLE 1 : Questions sur l'article Hurlin et al. "Machine Learning et nouvelles sources de données pour le scoring de crédit"

1. Résumer l'article en une demi-page
2. Qu'est-ce que c'est l'échantillon "out of sample" (page p5 du pdf) ? quelle est l'utilité de cet échantillon ?
3. Décrivez le lien entre le seuil et la matrice de confusion dans un processus de classification ?
4. En page 6, une illustration montre une courbe ROC. Quelle est la différence entre les 2 "mesures" de performance d'un modèle : l'accuracy (la justesse) et l'AUC de la ROC (l'aire sous la courbe ROC) ?
5. L'auteur parle d'"approches paramétriques standards". Qu'est ce qu'une approche paramétrique en apprentissage (machine learning) ? citez en une ? qu'est ce qu'une approche non paramétrique ? citez en une ?

ARTICLE 2 : Questions sur l'article Khiali et al. "Analyse des dynamiques spatio-temporelles à partir de séries temporelles d'images satellitaires".

1. Résumer l'article en une demi-page
2. Qu'est-ce qu'un graphe d'évolution ?
3. Qu'est-ce que qu'un synopsis et comment est-il utilisé pour l'approche de clustering de graphes d'évolutions ?
4. Comment l'approche proposée est évaluée ? Détailler le protocole expérimental ?
5. Quels sont les résultats de l'analyse conduite en Section 5.3 ?

ARTICLE 3 : Questions sur l'article Visoughi et al. "The spread of true and false news online"

1. Résumer l'article en précisant les principales contributions (une demi-page max).
2. Donner la définition de la notion de "cascade" et ses propriétés. Quel est le lien entre cette notion et la véracité d'une information, d'après les observations des auteurs du papier ?
3. Expliquer les statistiques considérées dans l'analyse et leur rôle (Fig. 4: A à F).
4. Donner les limitations de l'analyse comme identifiées par les auteurs.
5. Formuler 2 ou 3 critiques supplémentaires des travaux réalisés. Justifier la réponse.

ARTICLE 4 : Questions sur l'article "Extraction de Bases pour les Règles d'Association à partir des Itemsets Fermés Fréquents"

1. Résumer l'article en précisant les principales contributions (une demi-page max).
2. Préciser quelle est la différence entre des itemsets clos et des itemsets maximaux.
3. A votre avis, est-ce que le nombre d'itemsets maximaux est supérieur ou inférieur au nombre d'itemsets clos ? Justifier la réponse.
4. Est il possible à partir des itemsets maximaux d'obtenir l'ensemble des itemsets clos ? Justifier la réponse.
5. Dans le cadre de votre projet de classification vous avez utilisé différents traitements. Préciser comment l'utilisation de règles d'association pourrait aider à améliorer les classifieurs. Indications : les items peuvent être des mots pré-traités et la classe est connue.