

Examen HMIN232

session 1, 2021

- travail individuel
- dépôt sur Moodle le 17 mai 2021 **au plus tard**
- nom du fichier pdf: nom de l'étudiant + le numéro de l'article

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

Article 1: Cruz et al. Dictionary-Based Sentiment Analysis Applied to a Specific Domain

- 1 - Résumer l'article en une demi page.
- 2 - Expliquer la différence entre TrueSkill et AcroDefMI3. Donner un exemple de l'application des deux mesures.
- 3 - Les expérimentations ont été menées sur différents jeux de données. Préciser pour quel domaine l'approche est la plus efficace et donner les raisons possibles de la différence d'efficacité.
- 4 - La proposition permet de pouvoir extraire des opinions concernant des sujets particuliers mais n'est pas capable d'extraire le vocabulaire qui peut être associé. Par exemple s'il s'agit du domaine film, nous aimerions pouvoir apprendre automatiquement les termes acteurs, actrices, cinéma, etc. En vous inspirant de l'article, proposez une solution possible. Attention aux mesures.
- 5 - Donner trois faiblesses de l'étude. Proposez trois idées d'amélioration de l'approche.

Article 2: Wang et al. AN OPTIMAL SVM-BASED TEXT CLASSIFICATION ALGORITHM

- 1 - Résumer l'article en une demi page.
- 2.- Les auteurs adoptent une méthode de sélection de variables basée sur l'entropie. Expliquer cette méthode et justifier son choix dans l'étude. De manière générale, quel est l'objectif de la tâche de sélection de variables ?
3. - Quel schéma de vectorisation des données textuelles a été choisi dans cet article et pourquoi ? Donner deux autres possibilités de pondérer les documents textuels dans le cadre de machine learning.
4. - Expliquer le rôle du paramètre C pour les SVM et la méthode pour son choix élaborée dans l'article. Quels autres paramètres des SVMs connaissez-vous ? En nommer au moins un.
- 5 - Donner trois faiblesses de l'étude. Proposez trois idées d'amélioration de l'approche.

Article 3: Bordogna et al. Fuzzy Core DBScan Clustering Algorithm

- 1 - Résumer l'article en une demi page.
- 2 - Expliquer les motivations/nécessités liées à l'introduction d'une variante "fuzzy" de l'algorithme DBSCAN
- 3 - Quelle est la différence principale entre l'algorithme proposé et la méthode DBScan originale?
- 4 - Quelles sont les propriétés qui sont évaluées dans l'évaluation expérimentale? Détailler le protocole expérimental employé
- 5 - Donner trois faiblesses de l'étude. Proposez trois idées d'amélioration de la méthode Fuzzy Core DBScan

Article 4: Bouchet et al. Machine learning en Finances — [LIRE Jusqu'au chapitre 3 \(page 23 dans le livre, page 26 dans le pdf\)](#)

- 1 - Résumer l'article en une demi page.
- 2- Expliquez le lien entre ML et les statistiques
- 3- Expliquez ce qu'est une courbe de ROC ? Détailler comment on la trace ? Quelle est son interprétation

4- Quelles sont les limites du Machine Learning / Deep Learning "à tout va, en remplacement des acteurs" dans les cas d'utilisations, ici financiers ?

5 - Proposez trois idées d'amélioration sur les limites mises en évidence à la question 4.