

Compte-rendu

Présents : Sophie LÈBRE, Thomas AYRIVIÉ, Mehdi BELKHITER, Jamila CHERKAOUI, Dina EL HIJJAWI, Magatte LO.

Prochain rendez-vous : Vendredi 17 Novembre 2023 9h30, Université Paul Valéry, bâtiment B.

Travail à effectuer :

- Répartition des tâches et analyse exploratoire des données comme expliqué ci-après.
- Familiarisation du package R GLMNET.

Points abordés :

- 1) Définition date/heure du prochain rendez-vous.
- 2) Revue de la matrice.

On a utilisé Spark de Python, cette étape est effectuée, mais quantifier le nombre de NULL.

- 3) Revue de notre planification, étapes par étapes avec attributions.

Tâches réparties dans le rapport du 2e rapport intermédiaire.

matrice X : analyse descriptive des NAN ou NULL

=> remplacer par $2 \cdot 10^{-3}$

=> score : $-\log_{10}(\text{p-valeur})$

- **matrice Y** : analyse descriptive

=> construire les classes avec seuils proposés 2 et 1/2

- lien entre les variables de X :
 - à l'intérieur de la matrice X
 - corrélation 2 à 2 => représentation en heatmap

- conditionnement XtX : ratio des valeurs propres
 - ACP, clustering (première approche classification hiérarchique...)
-
- lien entre les colonnes de X et chaque colonne de Y : debut classif
 - régression logistique : essayer de coder un forward
 - découvrir glmnet et idéalement tester sur les données
 - PRÉVOIR des critères d'évaluation du modèle : précision et rappel,
-
- 4) Quelles visualisations recommandées dans l'analyse des données ? Et plus tard au semestre prochain ?
- En plus des représentations habituelles, représentation sous forme de heatmap, cases colorées.
-
- 5) Attendus de la commanditaire vis-à-vis du Pitch de la gestion de projet. Explication de notre approche de Pitch.
- Mme Lèbre sera au jury. Elle nous conseille de ne parler par que du thème mais aussi des analyses.