
EVALUATION

- ☞ L'évaluation du module se fait à travers un projet à réaliser en groupe.
- ☞ Chaque groupe comporte 3 étudiants. **Le nombre de sujets étant limité, aucun groupe de deux étudiants ne sera accepté.**
- ☞ Les sujets des différents projets sont de nature assez variée : il peut s'agir de lire et présenter les résultats d'un ou plusieurs articles de recherche, d'implémenter un algorithme vu en cours ou décrit dans un article de recherche, d'analyser un jeu de données, d'imaginer une méthode statistique permettant de répondre à une problématique donnée...
- ☞ Dans tous les cas, faire preuve d'initiative sera récompensé. Par exemple, si vous devez commenter un article, vous pouvez aller lire d'autres articles sur le sujet, ou bien télécharger des données présentées dans cet article et essayer de faire tourner un algorithme dessus...
- ☞ Les sujets de mémoire sont disponibles dans le fichier à télécharger
<http://alquier.ensae.net/sms-sujets.zip>
- ☞ Classez au moins 10 sujets par ordre de préférence et envoyer vos choix avant le mardi 18 avril à 18h à Pierre Alquier : alquier@ensae.fr
Exemple de choix :
Groupe constitué de ***, *** et ***
Nos souhaits:
1) Hebiri sujet 1
2) Lebarbier sujet 1
3) Carel sujet 2
4) Lebarbier sujet 3
etc...
- ☞ Lorsque j'aurai reçu tous les souhaits, je maximiserai l'utilité globale et vous communiquerai ensuite le numéro du sujet qui vous a été attribué.
- ☞ Les mémoires doivent être envoyés par mail à Pierre Alquier avant le mardi 16 mai 12h dernier délai, tout retard sera pénalisé.

Très brève description des sujets par intervenant (le fichier zip à télécharger contient une description plus détaillée de chaque sujet) :

— **Amandine Pierrot**

- 1) Analyse de données (fournies) de la compétition compétition GEFCON'14 par des modèles vus en cours et application à la prévision de la consommation d'électricité.
- 2) Sujet similaire mais pour la prévision des prix de l'électricité.

— **Charles Bouveyron**

- 1) Lecture de deux articles sur des méthodes statistique d'analyse des réseaux, et application à un jeu de données sur la blogosphère politique française.

- 2) Sujet similaire, avec deux articles différents portant sur une approche statistique légèrement différente.
- **Clara Champagne**
 - 1) résumer et commenter l'article de King *et al* sur la modélisation de la transmission du choléra.
 - 2) résumer et commenter l'article de Lourenço *et al* sur la modélisation de la transmission de la dengue.
 - **Edouard Oyallon**
 - 1) implémentation de la méthode de l'article Oyallon *et al* pour la classification d'images et application sur le jeu de données CIFAR10.
 - 2) implémentation de la méthode de l'article Oyallon *et al* pour la classification d'images et application sur le jeu de données STL10
 - 3) résumer et commenter l'article de Bruna et Mallat (2013) (*Invariant Scattering Convolution Networks*).
 - **Elie Guardiasabaoun**
 - 1) résumer et commenter l'article Blitzer *et al* sur l'analyse de sentiments.
 - 2) résumer et commenter l'article Daumé *et al* sur l'adaptation de domaine (NLP).
 - **Emilie Lebarbier**
 - 1) résumer et discuter un article sur une méthode de segmentation pour l'analyse des données en génomique : Robert Maidstone, Toby Hocking, Guillem Rigai and Paul Fearnhead (2014). On Optimal Multiple Changepoint Algorithms for Large Data. Preprint arXiv.
 - 2) idem, avec l'article : Jeremy J. Shen and Nancy R. Zhang (2012). Change-Point model on nonhomogeneous Poisson processes with application in copy number profiling by next-generation DNA sequencing. The Annals of Statistics.
 - 3) Implémentation d'une méthode de segmentation en génomique vue en cours (les slides sont fournis) et test sur un jeu de données (également fourni).
 - **Erwan Le Pennec**
 - 1) résumer et commenter l'article fondateur de Shannon sur la théorie de l'information.
 - 2) résumer et commenter un article sur les outils mathématiques utilisés dans la compression des fichiers audio MP3.
 - **Léna Carel**
 - 1) résumer et commenter l'article de Bouveyron *et al* sur le clustering des stations Velib.
 - 2) résumer et commenter l'article de El Mahrsi *et al* sur le clustering de passagers à l'aide de données billettiques.
 - **Mohamed Hebiri**
 - 1) résumer et discuter l'article : Florian Ziel, Rick Steinert, Sven Husmann (2015). Efficient modeling and forecasting of electricity spot prices. Energy Economics.
 - 2) résumer et discuter l'article : Johannes Lederer, Christian Müller (2014). Don't fall for tuning parameters : Tuning-free variable selection in high dimensions with the TREX. Preprint arXiv.
 - **Stéphanie Combes**
 - 1) analyse descriptive des réponses à la question "Quels sont les critères ayant conduit à la sélection de ce candidat ?" (enquête OFER de la DARES) : nuages de mots, clustering des individus...

- 2) construction d'un chatbot "conseiller d'orientation" à partir d'une base de données d'offres d'emploi scrappées sur le site internet Le bon Coin (60000 offres).