

# Pratique de la data science

April 11, 2025

## 1 Projet Final

**Objectif :** Mettre en place un pipeline permettant d'aggréger quotidiennement l'ensemble des signaux prédits par des modèles de clustering, classification, régression et d'étude de texte afin de fournir des recommandations pour la prise de décision sur le marché des actions.

**Consigne :** Le projet final est une combinaison de tous les TPs effectués durant le cours.

**Chaque groupe doit créer un Repo Github contenant :**

- L'ensemble des TPs en **format .py**.
- Un fichier **main.py** qui permet, à chaque exécution de :
  - Executer l'ensemble des fonctions de chaque fichier .py pour chaque entreprise.
  - Mettre en place une stratégie d'aggrégation des différents signaux pour fournir des recommandations.
- Un rapport de 10 à 12 pages permettant d'expliquer les motivations, la démarche scientifique, les résultats et la mise en place d'une stratégie d'investissement.

### 1.1 Pipeline

Le fichier main.py doit permettre de lancer automatiquement (chaque jour) une analyse complète et actualisée des entreprises étudiées.

L'objectif est de produire un fichier de sortie contenant, **pour chaque entreprise :**

- Un conseil d'investissement.
- Une liste des entreprises similaires (clustering).
- Une prévisions de rendement à court terme.
- Les titres de news du jour et une analyse de sentiment globale (qui pourra être aggrégée aux prédictions pour le conseil d'investissement).

### 1.2 Contenu du rapport

Le rapport final doit être présenté comme un papier de recherche (rédigé en latex). L'objectif est de documenter les motivations et les différentes méthodes utilisées ainsi que les résultats obtenus et les interprétations.

**Exemple de sections :**

1. Introduction / Motivations
2. Related works (Courtes présentations de 2/3 papiers qui traitent du sujet).
3. Clustering
4. Classification "buy", "sell", "hold"

5. Prédiction de rendement à  $J+1$
6. Analyse de sentiments sur News financières et
7. Stratégie d'aggrégation pour fournir des recommandations pertinentes.
8. Conclusion

Pour les parties 3 à 6, il faudra **préciser** :

- Objectif.
- Méthodes utilisées (ML et DL) .
- Résultats et interprétations.

### 1.3 Modalité de notation et rendu

L'évaluation est faite sur la qualité du rapport et du code. Tout ajout par rapport aux TPs (**nouvelles méthodes, sources de données, types de signaux...**) sera fortement valorisé.

Le projet final est à déposer sur un repo GitHub partagé entre les 3 membres du groupe.

Le lien du repo doit être envoyé par mail à pierre.fihey@dauphine.eu avant le **18 mai à 23h59**.