



DATA MINING

Introduction générale



Energizer Balle aux mots

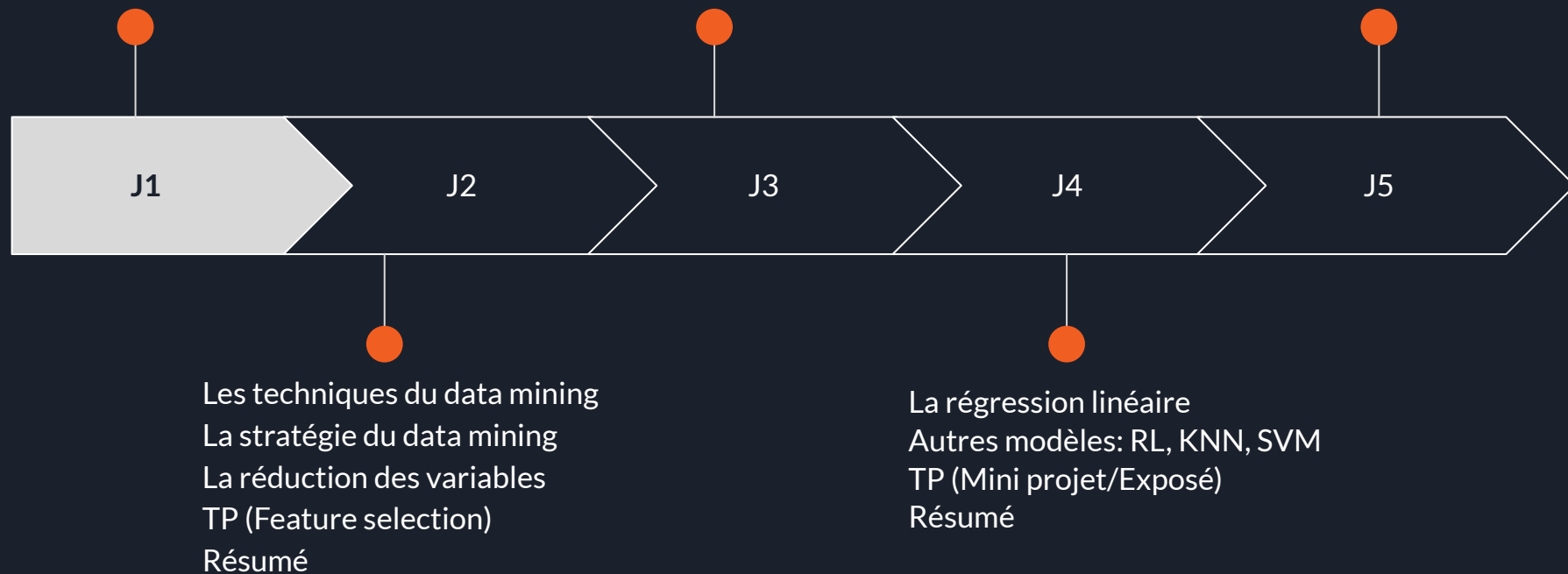
Description de l'activité:

Je lance une balle et la personne qui la reçoit doit se présenter et donner quelques informations supplémentaires sur elle (résolution 2023, activités en entreprise etc.). Ensuite, elle relance la balle à une autre personne qui se présente à son tour.

Organisation du cours
Introduction générale
TP (Prise main des outils de travail)
Résumé

La modélisation décisionnelle
Les méthodes factorielles
TP (ACP & AFD)
Résumé

Révision générale
Evaluation
Bonus





La note finale de l'enseignement

- Partie pratique 70% (14 points)
- Partie théorique 30% (6 points)
- 1,25 points de bonus:
 - Présence && 100% de bonnes réponses
 - Activités notées pendant la séance (à raison de 0,25 par séance)



Exposés

Travail demandé:

- Présentation
 - Problématique | Dataset
- Démo
 - Analyse exploratoire | Explications
 - Prédiction

Liste des sujets non exhaustive:

- Text mining (Word2vec, Spacy, Transformers...)
- Les méthodes d'évaluation d'un modèle d'apprentissage (Validation croisée, GridSearch...)
- Application du modèle RandomForest
- Application du modèle SVM (SVC...)



Contenu

- Introduction générale
 - Définitions (Data mining, IA, Data Science, Données, Informations...)
 - Motivations
- Travaux pratiques
 - TP1: Prise en main de Python
 - TP2: Prise en main de Pandas et Numpy
 - TP3: Prise en main de Matplotlib et Seaborn
- Résumé

Quizz / Mots croisés

Introduction générale → Définitions

Data mining

- Depuis toujours, et surtout depuis la naissance des CRM (peut-être avant même), le besoin ultime des directeurs marketing est la connaissance client (comportement du client et ses envies).
- Ensemble de techniques d'analyse statistique des données (ou analyse exploratoire des données) qui permettent d'assurer une tâche bien précise.
- L'objectif principal est l'extraction de la connaissance (ou bien de l'information utile) à travers l'étude des relations entre les différentes features de la DATA.

Introduction générale → Définitions

Intelligence artificielle

- C'est l'ensemble des techniques qui permettent de produire un système intelligent.
- Objectif: Simuler l'intelligence humaine
- Informations diverses:
- IA responsable, les réglementations, France, commission européenne
 - CNIL:
 - RGPD
 - DGA (Data Governance Act) adopté en mai 2022 et applicable en septembre 2023
 - Objectif: création des structures
 - Plein d'autres lois: DA, DMA, DSA, EHDS
 - [Stratégie européenne pour la donnée : la CNIL et ses homologues se prononcent sur le Data Governance Act et le Data Act | CNIL](#)

Introduction générale → Définitions

Machine learning

- Techniques qui utilisent des algorithmes qui sont capables d'apprendre tout seuls, à travers l'usage d'un grand volume de données.
- Data driven, auto-apprentissage, auto-amélioration (MLOps)
- Dataset: Train / Test
- Objectif: prendre des décisions (prédictions, recommandations) à travers le modèle développé.
- Exemples:
 - diagnostics médicales
 - détection de fraudes
 - campagnes publicitaires/marketing

Introduction générale → Définitions

Data Science

- La data science vise à résoudre des problèmes complexes en utilisant une approche interdisciplinaire qui combine des connaissances en informatique, en statistiques et en domaines métier spécifiques.

Introduction générale → Définitions

Donnée

- Donnée ou data, c'est la valeur qui correspond à un champ donnée ou une variable donnée.
- Types:
 - Données discrètes: binaires, énumératives
 - Données continues: entières, réelles (âge, salaire...)
 - Dates
 - Données textuelles
 - Multimédia: Images, Vidéos, Sons...
- Natures:
 - Variables quantitatives
 - Variables qualitatives (ou catégorielles)

Introduction générale → Définitions

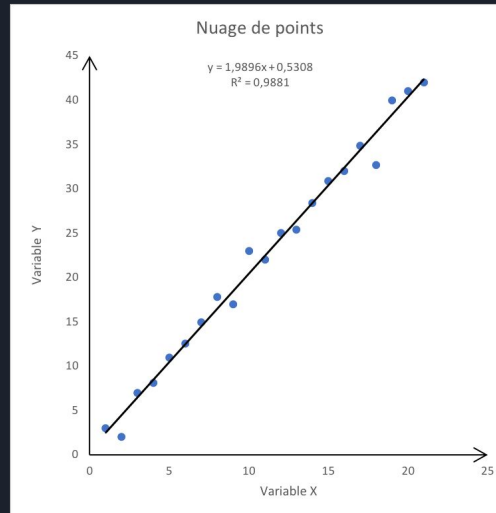
Information

- Une information est une donnée utile.
- Notre objectif est de collecter le maximum possible d'informations à travers le tas de données qu'on a afin de découvrir les relations et les patterns de la data.

Introduction générale → Définitions

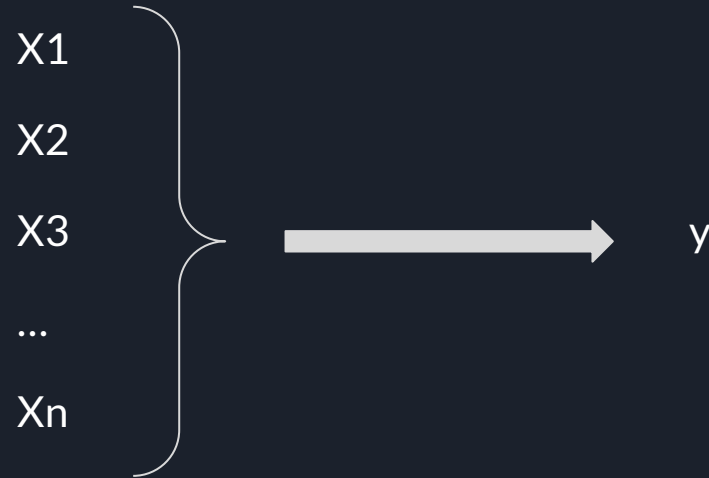
Modèle

- Un modèle est une fonction donnée qui a un objectif bien déterminé.
- Exemple: $f(x) = ax+b$ est un modèle linéaire.



Introduction générale → Définitions

- Feature (ou variable explicative)
- Target (ou variable expliquée)



Introduction générale → Motivations

- **Pourquoi:**
 - Cloud Computing
 - IoT et les objets connectés
 - BiG DATA
 - Maîtrise de l'avenir
 - Tendances
- **Qui:**
 - Energie
 - Banque
 - Défense
 - Marketing
 - Télécom
 - etc.
- **et après ?**
 - Aide à la décision
 - Supervision claire: métriques, objectifs mesurables
 - Gain: Pertinence, Temps, Efficacité

TP n°1

Prise en main de Python

TP n°2

Prise en main de Pandas et Numpy

TP n°3

Prise en main de Matplotlib & Seaborn



Résumé

Qu'est-ce qu'on a appris aujourd'hui ? (5 minutes de réflexion)





Références

- [Data mining I Exploration Statistique](#)
- [Data Mining et Statistique](#)
- [5 outils data mining pour mieux analyser vos données - Codeur Blog](#)
- [Data Mining définition : Qu'est-ce que l'exploration des données ?](#)
- [Difference in Data Mining Vs Machine Learning Vs Artificial Intelligence](#)



Merci 😊

formateur_mohamed.zwawa@supdevinci-edu.fr