

Projet Final : Machine Learning & Deep Learning

Exploration approfondie d'un problème réel de Data Science

K. Kadri

- 1 Contexte général
- 2 Objectifs pédagogiques
- 3 Tuning & Explicabilité
- 4 Livrables
- 5 Critères d'évaluation

Objectif du projet

- Mener un projet complet de Machine Learning (et un peu de Deep Learning).
- Étudier un dataset réel, choisi librement par l'étudiant.
- Explorer un angle original ou une problématique approfondie.

Attention

Le but n'est pas seulement d'obtenir une bonne accuracy, mais de conduire une **démarche complète de Data Scientist**.

Compétences visées

- Prétraitement, Feature Engineering, Feature Selection.
- Entraîner plusieurs modèles ML différents.
- Tuning d'hyperparamètres.
- Explicabilité (SHAP, importances).
- Construction d'un modèle DL simple (ANN ou CNN).
- Analyse critique, limites et conclusion.

- Identifier la structure, la distribution, les outliers.
- Définir clairement le problème : classification, régression, clustering.

Question clé

Quelle est la vraie nature du problème que vous tentez de résoudre ?

Pipeline ML complet

- Feature Engineering
- Standardisation / Normalisation
- Encodage catégoriel
- Sélection de features
- Cross-validation

Bonne pratique

Utiliser un pipeline clair, reproductible et bien justifié.

Modèles ML obligatoires

- Régression (linéaire ou logistique selon dataset)
- KNN
- Decision Tree
- Random Forest
- SVM
- Comparaison structurée entre modèles

- ANN (données tabulaires) **ou**
- CNN (images)
- Analyse des courbes de loss : overfitting, régularisation

Conseil

Garder un modèle simple mais bien expliqué.

Hyperparameter Tuning

- GridSearch
- RandomSearch
- (Optionnel) Optuna / Optimisation bayésienne

Attendu

Comparer avant/après tuning pour au moins un modèle.

- SHAP Summary Plot
- SHAP Local (Force Plot)
- Feature Importances (RF ou Gradient Boosting)

Compétence clé

Savoir interpréter les résultats, pas seulement les calculer.

Structure du rapport final

- ① Présentation du problème
- ② Description du dataset
- ③ Analyse exploratoire (EDA)
- ④ Feature Engineering
- ⑤ Modèles ML testés
- ⑥ Tuning des hyperparamètres
- ⑦ Explicabilité (SHAP)
- ⑧ Modèle DL
- ⑨ Conclusion + limites

Essentiel

Le projet doit être clair, reproductible, structuré et argumenté.

- Code propre (Notebook + Python).
- Expérimentations rigoureuses.
- Visualisations pertinentes.
- Interprétation rigoureuse.
- Approche critique des limites.

A — Rigueur technique (40%)

- Pipeline ML propre et reproductible.
- Bon usage des algorithmes.
- Preprocessing cohérent.
- Modèle DL fonctionnel (même simple).

B — Feature Engineering (20%)

- Création intelligente de nouvelles features.
- Sélection cohérente (statistique ou model-based).
- Impact sur les performances justifié.

Note

Un projet avec peu de features mais bien travaillées vaut plus qu'un modèle complexe mal expliqué.

C — Explicabilité (20%)

- Utilisation correcte de SHAP.
- Interprétation des variables clés.
- Analyse des erreurs.

D — Exploration approfondie (20%)

- Approche créative de la problématique.
- Analyse poussée : biais, métriques avancées, seuils, résidus.
- Validation croisée rigoureuse.
- Initiative personnelle au-delà de l'automatisation.

Critère premium

Ce qui compte n'est pas l'accuracy brute, mais votre **profondeur d'analyse et votre compréhension**.