

FICHE MINI-PROJET : Business Intelligence avec Python

Objectifs du Mini-Projet

Maîtriser l'ensemble du pipeline BI : Collecte → Préparation → Analyse → Visualisation → Déploiement

- Comparer plusieurs datasets pour identifier le plus performant
- Développer une interface Streamlit interactive pour tester les modèles
- Bonus : Déploiement (Heroku/Render), datasets supplémentaires (+ points)

Étapes Obligatoires du Projet

1. Choix & Comparaison des Datasets (15 pts)

Analyser 2+ datasets (structure, qualité, taille, valeurs manquantes)

Créer tableau comparatif : features, target, performance baseline

Justifier le choix du dataset principal

Documenter les sources (Kaggle, UCI ML, etc.)

2. Prétraitement des Données (20 pts)

Pandas + NumPy :

Nettoyage (missing values, outliers, doublons)

Encodage catégorielles (OneHot, Label, Target)

Normalisation/Scaling (StandardScaler, MinMaxScaler)

Feature Engineering (création de nouvelles variables)

Gestion des classes déséquilibrées (si applicable)

3. Modélisation (25 pts)

Scikit-learn (minimum 3 modèles) :

Logistic Regression, Random Forest, XGBoost, SVM, Gradient Boosting

Train/Test split (80/20 ou 70/30)

Métriques : Accuracy, Precision, Recall, F1-Score, ROC-AUC

Comparaison performances entre datasets et modèles

4. Visualisation & Analyse (15 pts)

Matplotlib + Plotly :

Exploratory Data Analysis (EDA) : histogrammes, distributions, boxplots

Matrice de corrélation

Feature importance (Random Forest, XGBoost)

Courbes ROC et AUC

Matrice de confusion

Comparaison performances modèles/datasets

Insights et recommandations

5. Interface Streamlit (25 pts)

Application web interactive

Layout professionnel avec tabs/sidebars

Upload fichier CSV ou dataset pré-chargé

Prévisualisation et statistiques données

Sélection du modèle et hyperparamètres

Prédictions en temps réel sur nouveaux exemples

Visualisations dynamiques (graphes interactifs)

Affichage métriques et performance

Export résultats (CSV/JSON)

Options Bonus (+ points supplémentaires)

Déploiement Web :+10 pts Héberger sur Streamlit Cloud / Heroku / Render

Dataset Supplémentaire +5 pts/dataset Analyser 3+ datasets (max 15 pts)

Livrables Requis

1. Notebook Jupyter (.ipynb) :

Code complet ,Exécution fluide sans erreurs,Visualisations intégrées , Explications des décisions

2. Application Streamlit (.py) :

Démarrage : streamlit run app.py

Interface intuitive et fonctionnelle

Gestion d'erreurs robuste

3. Rapport PDF (5-10 pages)

Résumé exécution

Comparaison des datasets

Méthodologie et choix

Résultats et performance modèles

Screenshots interface Streamlit

Conclusions et recommandations

Lien déploiement (si applicable)

4. Vidéo Démonstration (5 minutes)

Lancer l'application

Charger/explorer données

Effectuer prédictions

Afficher visualisations

(Optionnel : montrer déploiement)

Résultats non probants

Date limite : 20/12/2025

Dépôt dans drive : Code + PDF rapport + Lien Streamlit (le cas échéant) + Vidéo

Mail : ai.bensalem@esi-sba.dz