Python for Data Science 2: Guided Machine Learning Project

University: Tek-Up
Group: ING-4-SDIA

Tutor: Haythem Ghazouani

Academic Year 2025-2026

1 Project Overview

This module is structured as a 100% project-based learning experience over 13 weeks. Students will follow a guided pipeline to build a reproducible and deployable Machine Learning solution. The focus is on advanced concepts including Ensembling, Boosting, Feature Selection, and MLOps.

2 Project Synopsis & Dataset Selection

The **Bank Customer Churn** example used in these tutorials serves as a **Project Synopsis** (a baseline example).

Important

Each team is free to choose their own objectives and dataset (either ready-to-use from repositories like Kaggle/UCI or self-scraped). However, the choice of dataset and objective **must be validated** by the tutor before proceeding.

- **Target variable:** Churn (Binary: 0 or 1).
- **Challenges:** Mixed data types, class imbalance, and feature engineering opportunities.

3 13-Week Roadmap

4 Project Structure

The workspace is organized as follows:

- `cours/`: Theoretical supports and slides.
- `code/`: Structured Python scripts (GitHub-ready).

Week	Phase	Key Deliverable
1	Setup	Environment & Git Init
2	Step 0	Web Scraping (Optional) / Data Acquisition
3	Step 1	Exploratory Data Analysis (EDA)
4	Step 2	Data Preprocessing Pipeline
5	Step 2	Feature selection & Engineering
6	Step 2	Modeling: Ensembling (RF, Voting)
7	Step 2	Modeling: Boosting (XGBoost/LightGBM)
8	Step 2	Experiment Tracking with MLflow
9	Step 3	API Development with FastAPI (Part 1)
10	Step 3	API Development with FastAPI (Part 2)
11	Step 4	Frontend Development with React
12	Step 5	Containerization with Docker
13	Step 5	Deployment & Final Review

Table 1: Project Planning Schedule

- `tutos/`: LaTeX tutorials for each step.
- `data/`: Dataset storage (raw and processed).

5 Deliverables

Students are expected to submit a link to their GitHub repository containing:

1. A complete ML pipeline tracked with MLflow.
2. A functional FastAPI backend.
3. A React frontend for model inference.
4. Docker configuration files for full-stack deployment.