# Compact Screening Project (4 – 8 hrs)

|  |  |
| --- | --- |
| **Aspect** | **Spec** |
| **Problem** | **Bank-Marketing Lead Conversion** – predict whether a prospect will subscribe to a term-deposit product after a phone campaign. |
| **Dataset** | “Bank Marketing” (Kaggle). 45 421 rows · 16 features (age, job, balance, contact type, previous outcome, etc.) plus y (yes/no).  Link to read more about the dataset: <https://www.kaggle.com/datasets/willianoliveiragibin/bank-marketing?resource=download> |

# Tasks

1. **Exploratory Data Analysis (EDA)**  
   *Summarize class balance, highlight two interesting feature–target relationships (plots optional).*
2. **Pre-processing pipeline**
   * Handle missing values (justify method).
   * Encode categorical variables (one-hot or target encoding).
   * Train/validation split or k-fold.
3. **Baseline model**  
   Train **any** classic algorithm (e.g., Logistic Regression, Random Forest, XGBoost). Report **at least two** metrics: *accuracy* **and** *AUC or F1*.
4. **Improvement step**  
   Try one of: feature engineering, hyper-parameter tuning, class-imbalance technique, or calibrating probabilities. Briefly justify the choice.
5. **Brief evaluation**  
   ½-page summary: what worked, what didn’t, next steps if you had more time.

# Expected deliverables

|  |  |
| --- | --- |
| **File** | **Contents** |
| bank\_marketing.ipynb or .py | Clean, runnable code with section headings. |
| README.md | Setup instructions (≤5 lines) plus the ½-page summary. |
| *(optional)* requirements.txt | Only if non-standard libs used. |

# Scoring rubric (100 pts)

|  |  |  |
| --- | --- | --- |
| **Criterion** | **Pts** | **What we look for** |
| Correct loading & basic EDA | 15 | Can they spot class imbalance / obvious quirks? |
| Clean preprocessing pipeline | 20 | Reproducible, no data leakage. |
| Baseline model & metrics | 25 | Reasonable choice, metrics computed on hold-out data. |
| Improvement attempt | 15 | Any thoughtful step that beats the baseline. |
| Code quality & organization | 15 | Readable, commented, modular. |
| Communication (README) | 10 | Clear, concise, no jargon overload. |