The goal of this project is to build data-driven models that predict the success of telemarketing calls aimed at selling long-term bank deposits. Data mining extracts actionable patterns from large datasets; in this context, it can reveal customer and campaign factors that meaningfully improve business decisions.

Direct marketing typically targets specific customer segments. Selecting the best set of clients—i.e., those most likely to subscribe—is known to be NP-hard. Your task is to develop models that automatically identify and prioritise customers to call, supporting managers in scheduling the next best contacts during bank marketing campaigns.

We will use a dataset collected by a retail bank between 2008 and 2013, a period spanning the global financial crisis of 2008. This setting provides a realistic, high-variance environment for training and evaluating predictive models.

Using the algorithms covered in class, your goal is to build predictive models from the provided dataset. The project must follow the CRISP-DM methodology and include Python code for each phase:

1. business understanding and data understanding,
2. data preparation (cleaning, transformation, feature engineering),
3. modelling (training and tuning), and
4. evaluation.

Submit a written report that documents, in detail, the process you followed. The report must include:

- **Data mining goals** (problem statement, target variable, success criteria).
- **Exploratory analysis**: key plots/tables and their interpretation.
- **Data cleaning & preprocessing**: what you did and why (handling missing values, encoding, scaling, outliers, class imbalance, etc.).
- **Modelling**: algorithms tried, hyperparameter choices, validation strategy.
- **Evaluation & interpretation**: results, error analysis, limitations, and assumptions/constraints made during development.
- **Conclusions & recommendations**: proposed policies or actions informed by your findings.

Assessment

The primary basis for grading is the **quality and rigour of your analysis process** (clarity, justification, reproducibility), followed by the **soundness of conclusions** and **practical recommendations**. Model accuracy matters, but **your reasoning, methodology, and interpretation carry more weight than raw performance scores**.

Deadline and submission instructions

- The project should be submitted to Moodle (course page) by **23:59 on Sunday, 26 October 2025**.
- Submissions received until 23:59 on Monday, 27 October 2025 incur a **10% grade penalty**.
- After 27 October 2025: Not accepted (grade = 0).

What to submit

- A single **ZIP** containing all code and the report.
- Filename format: MINDD-XXX-NumberX-NumberY.zip
  - XXX = your PL teacher's acronym (MFC – Fátima Rodrigues; AZC – Catarina Figueiredo).
  - NumberX, NumberY = the student numbers of each group member (add more numbers if your group has >2 members).
- Only **one member group** submits the project on Moodle.

Presentations and grading

- The PL teacher will assess each group during the next two weeks, according to the previous schedule. All group members' attendance is mandatory.