Data Mining Project

- **Deadline:** 31 May 2025
- **Submission Folder (Rar/Zip) must include:**
  - **Presentation** (summary slides for oral defense)
  - **Final Report** *(maximum 10 pages)*
  - **Notebook** containing all code and data mining pipeline
  - **README file** with: Full names of engineering students, ENSAM student codes
- **Group Composition:** 6 to 7 engineering students per group
- **Mail:** hosniteaching[at]gmail[dot]com

Project Themes and Datasets (choose one theme):

1. **Predictive Maintenance of Industrial Machines**
   - NASA Turbofan Engine Degradation Dataset (C-MAPSS)
   - PHM 2008 Challenge Dataset
   - Kaggle Predictive Maintenance Dataset
2. **Environmental Data Mining**
   - Air Quality Data Set (UCI Repository)
   - OpenAQ API (real-time global data)
   - Delhi PM2.5 Air Quality Dataset (Kaggle)
3. **Market and Customer Analysis**
   - Online Retail Dataset (UCI / Kaggle)
   - Telco Customer Churn Dataset (Kaggle)

Project Expectations:

You are expected to **implement the entire data mining pipeline**, including:

1. **Data Preprocessing**
2. **Exploratory Data Analysis**
3. **Feature Engineering**
4. **Model Selection and Training:** Use **multiple machine learning techniques**
5. **Evaluation of Models**
6. **Interpretation of Results**

Important Requirements:

- **At each step**, justify the choice of statistical or machine learning techniques used.
- Provide a **clear rationale** behind the methods and decisions taken.
- Include **a critical analysis of the results** in the report, highlighting insights, strengths, and limitations.