Energy

Data Science Project Project Statement

Project Overview

This project aims to apply data science methodologies and Python programming to analyze data and generate actionable knowledge. The overarching theme of the project is **Energy**.

Students are expected to formulate a clear research question within this domain and develop a complete data science study to address it.

Data Sources

Students may collect data from credible and authoritative sources, including, but not limited to:

- `https://e-redes.opendatasoft.com/`
- `https://datahub.ren.pt/`
- `https://precoscombustiveis.dgeg.gov.pt/estatistica/preco-medio-diario/`
- `https://www.erse.pt/en/home/`
- `https://climate.nasa.gov/`
- `https://www.ncei.noaa.gov/cdo-web/datasets`
- `https://climate.copernicus.eu`
- `https://ourworldindata.org/`
- `https://databank.worldbank.org/source/sustainable-development-goals-(sdgs)`
- INE
- Lisboa Aberta (`https://dados.cm-lisboa.pt/`)
- Dados Gov (`https://dados.gov.pt/pt/`)

Kaggle datasets are not allowed.

Other credible sources may be used upon validation. Web scraping is allowed. Data may also be requested from ISEG services when appropriate.

Project Requirements

- Each group must define a unique research question related to Energy.
- The proposed research question must be shared in the appropriate course Team channel.
- Duplicate research questions are not allowed.
- The project must follow the CRISP-DM methodology[1], as presented in class.
- Students are encouraged to apply multiple data science techniques, including:
  - Network science
  - Unsupervised learning algorithms
  - Supervised learning algorithms
  - Time series analysis
  - Natural Language Processing (NLP)
  - Data visualization
  - Data preparation, including pre-processing, and data cleaning
- The analysis may be conducted at the individual, corporate, regional, or country level.

Deliverables

The final submission must include:

- A fully functional Jupyter Notebook and/or Python application.
- Optionally, a small web application developed in Flask (not mandatory).
- A written report in PDF format (DOC/DOCX or LaTeX source files must also be submitted).

All materials must be accessible to the course lecturer.

Deadlines and Presentations

Deadline

Please refer to the official worksheet (Excel file) for submission deadlines.

Proposal Presentation

- Date: Verify in the excel File
- Duration: 5 minutes (expected)

Final Presentation

- Maximum duration: 15 minutes (10 minutes recommended)
- Dates: Verify in the excel file.
- All group members must participate in the presentation.
- Attendance is mandatory for all students.

Submission Platform

All materials must be submitted via the MS Teams group channel.

Report Structure

The report must follow a structure similar to a scientific research paper:

1. **Title**
2. **Abstract**
3. **Keywords**
4. **Introduction**
   Clearly state:
   - The research problem addressed.
   - The objectives of the study.
   - The contributions of the group.
5. **Literature Review**
   Identify and discuss prior studies related to the chosen topic.
6. **Methodology**
   The project must follow the CRISP-DM/POST-DS methodology [1], including:
   - Business Understanding
   - Data Understanding
   - Data Preparation
   - Modeling
   - Model Validation
   - Deployment
7. **Results**
   Present and clearly identify the main findings.
8. **Discussion**
9. **Conclusions**
10. **References**

References

[1] C. J. Costa and J. T. Aparicio, "POST-DS: A Methodology to Boost Data Science," *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, Seville, Spain, 2020, pp. 1–6. doi: 10.23919/CISTI49556.2020.9140932.