

# CS 132 Machine project

## Instructions

### 1. Goal

The main goal of the machine project is to assess students' knowledge and skills on Data Science in real-world scenario.

By following the five steps of Data Science, such as

- 1) Asking an interesting question,
- 2) Obtaining the data,
- 3) Exploring the data,
- 4) Modeling the data,
- 5) Communicating and visualizing the results,

students will work in a group setting towards answering their proposed research question.

### 1.1 Theme for the semester

For this semester, projects will be inspired by the theme: “*Climate Change: current state and actions*”.

---

## 2. Grouping

The machine project will be a group work. When forming groups:

- Groups must have 3-4 members
- Students from sections WFX and WFY may be in the same group

Email to [crrael@up.edu.ph](mailto:crrael@up.edu.ph) the name of the members of your group on or before April 1, Friday. A student who does not belong to a group by April 2 will be randomly assigned to a group.

---

### 3. General Criteria

Students' projects will be graded based on the following:

- **Research:** students must have a clear and concise research question for the project; *what is the research question you are trying understand more by extracting knowledge from data, why is this important, and who will benefit from this*
  - **Data processing:** students must be clear and specific in the data they collected; *why is their data appropriate for the problem; how do they preprocess the data*
  - **EDA:** students must show the exploratory analysis techniques they applied to understand the data: *what are the characteristics of the data that support your preprocessing and modeling decisions?*
  - **Model:** students must choose and implement a reasonable model / solution; *why did they choose the specific algorithm / model*
  - **Results and discussion:** students must be able to communicate and visualize their results, i.e., the answers to the research question; *how do they interpret the results; what is the conclusion; what are issues in the project that needs to be addressed in the future*
- 

### 4. Grading

The machine project is equivalent to 20% of the students' final grade.

The formula for grading the machine project is as follows:

$$\begin{aligned} \text{Final MP grade} = & \\ & 10\% * [\text{MP proposal grade}] * [\text{Proposal peer grade}] \\ & + 70\% * [\text{MP report grade}] * [\text{Coding peer grade}] \\ & + 20\% * [\text{MP poster grade}] * [\text{Poster peer grade}] \end{aligned}$$

#### 4.1 Proposal grade (10%)

**Deadline for proposal: April 29, 2022**

During the proposal stage, groups will propose their research question.

Their proposal will be submitted as a Wiki page in UVLE, and will include the following:

- Project summary
- Background
- Materials and methods (for data collection)
- References

Groups can continue updating their proposals until the submission deadline. Afterwards, their proposal will be graded as follows:

- 100% : The proposed project is **excellent**. The research question is clear, the proposed data collection method is appropriate, the project is relevant to the theme, and the target results are achievable.
- 50% : The proposed project is **acceptable**. Some minor revisions are needed, e.g. clarify the research question, update the data collection method.
- 25% : The proposed project is **similar to an already existing information/knowledge**. The group should revise the scope of the problem, type of data, or method to be used.
- 0% : The proposed project is **not acceptable**. Major revisions are needed.

#### 4.2 Report grade (70%)

The group's report should contain (but not limited to) the following:

- Codes/script
- Data source
- Data processing
- EDA/plots
- Modelling

**Deadline for report: 31 May**

#### 4.3 Poster grade (20%)

To Communicate and visualize your project's results, you will come up with a poster. The poster should convey what your research problem is about and your results in away that your target audience will easily understand.

**Presentation: Jun 1, 2022**