Deliverables

For each project, each team will need to hand in the following deliverables :

- A report summarizing the experiments and the results of the project
- The code written as part of the project. Notably to :
    - Generate, analyze, visualize and pre-process the data
    - Train and finetune models
    - Run a model on a data set to generate predictions
    - Analyse the results
- The best model obtained (this model will be used by the TAs for the blind test set evaluation)
- Each team will also be required to present their work in front of their peers. This presentation will summarize the experiments, results and observations for the project.

Finally, each student will be required to submit peer evaluations for everyone in his team (see section “Individual Grades” for more information). These peer evaluations will happen at two moments during each project :

- Halfway during the project, every student will hand in “informal” peer evaluations. These evaluations will serve to monitor how each team is doing. These evaluations will have **no** effect on any student’s grade. The TAs will report to each student only his **average** peer evaluation so that the student has an opportunity to adjust his contribution, if needed, before the end of the project.
- At the end of the project, every student will hand in the official peer evaluations. These evaluations will have an effect on the students’ final project grades.

Deadlines

Each team needs to provide the deliverables (report + code + best model) corresponding to a project at the latest on 11:59pm on the day before the project presentations. For every late day, the project will see its grade reduced by 10 percentage points (out of 100). After 4 late days, the project will automatically get 0%. Example for a project with the presentations on a Friday :

- Project handed in at the latest on Thursday 11:59pm : No penalty
- Project handed in before Friday 11:59pm : 10% penalty
- Project handed in before Saturday 11:59pm : 20% penalty
- Project handed in before Sunday 11:59pm : 30% penalty
- Project handed in before Monday 11:59pm : 40% penalty
- Project handed in after Monday 11:59pm : 100% penalty

The deadlines for the peer evaluations are as follows :

- Informal peer evaluations : Friday 11:59 pm on the third week of the project
- Official peer evaluations : Friday 11:59 pm on the sixth week of the project

Individual Grades

Though the projects are built in teams, each student will obtain an individual grade out of them. A student's individual grade for a project will be obtained by multiplying the overall grade for the group project (the same for every student in a team) with an individual adjustment factor which will reflect each student's personal contribution to the project as a whole. The process is derived from the method presented in this article with a few modifications.

At the end of a project, each team member will be asked to evaluate the contribution of the other team members. Each team member will have 100 "contribution" points to split between him/herself and the others. Example :

Eval Of		Eval by		
	Student A	Student B	Student C	Student D
Student A	25	27	30	25
Student B	23	25	30	25
Student C	47	46	36	25
Student D	5	2	4	25
SUM	**100**	**100**	**100**	**100**

Every student's contribution score is obtained by taking the average of the evaluations given to him/her by the **other** students (a student's own evaluation doesn't count). These average contribution scores are then rescaled such that, within a team, they sum to 100.

Eval Of		Eval by				
	Student A	Student B	Student C	Student D	Average	Rescaled
Student A	25	27	30	25	27,33	28,37
Student B	23	25	30	25	26,00	26,99
Student C	47	46	36	25	39,33	40,83
Student D	5	2	4	25	3,67	3,81
SUM	**100**	**100**	**100**	**100**	**96,33**	**100,00**

The final step is to transform the contribution scores into **bounded** adjustment factors. The bounds for the adjustment factors will be decided in class with the help of the students. Here we will use 85% - 115% as en example.

- The first number is the minimum possible adjustment factor and corresponds to the adjustment factor that a student with a contribution score of 12.5 (half of the expected contribution score if everyone contributed equally) or less would obtain.
- The second number is the maximum possible adjustment factor and corresponds to the adjustment factor that a student with a contribution score of 50 (twice the expected contribution score if everyone contributed equally) or more would obtain.

To convert the contribution scores to adjustment factors, we constrain all contribution scores to lie in the interval [12.5, 50.0] and then fit a quadratic polynomial to transform these bounded contribution scores to adjustment factor. A contribution score of 12.5 will correspond to an adjustment factor of 85% (in our example), a contribution score of 25 will correspond to an adjustment factor of 100%, and a contribution score of 50 will correspond to an adjustment factor of 115% (in our example). For other contribution scores between 12.5 and 50, we will use a quadratic polynomial, fitted on these three data points, to compute the adjustment factor.

Eval Of		Eval by							
	Student A	Student B	Student C	Student D	Average	Rescaled	Floor	Ceiling	Adjustment factor
Student A	25	27	30	25	27,33	28,37	28,37	28,37	103,19%
Student B	23	25	30	25	26,00	26,99	26,99	26,99	101,93%
Student C	47	46	36	25	39,33	40,83	40,83	40,83	111,82%
Student D	5	2	4	25	3,67	3,81	12,50	12,50	85,00%
SUM	**100**	**100**	**100**	**100**	**96,33**	**100,00**			

Very important : you are asked to base your peer evaluations not only on their contribution of results, but also on their contribution of time and contribution of effort. From the article :

1. ***Individual's time devoted to the project:*** *This includes time spent working alone or with other group members on delegated tasks, attending group meetings (whether in person or by Internet chat), and otherwise communicating with group members (telephone, e-mail, bulletin boards).*

2. ***Individual's quality of effort****: This aspect is intended to encompass "attitude" and the spirit of being a "team player." It distinguishes members who may dutifully come to every meeting but sit there saying nothing, from those who come prepared to show what they've done, to actively discuss and to plan. It distinguishes those who do nothing unless having been ordered and badgered, from those who sacrifice their personal convenience, eagerly look for tasks, "fight fires," and volunteer to do whatever is needed, especially in the face of looming deadlines.*

3. ***Individual's quality of result:*** *This is the objective evaluation of what a member produces (e.g., a written section of a design document, a software module, the final editing of a user's manual), regardless how much or how little time and effort were needed to produce it.*

The contribution scores you assign to your teammates should reflect your honest opinion. We reserve the right, if we suspect foul play within a team, to step in and review the scores. If this happens, we will notify the team in question.

Project evaluation (each project is worth 50% of the final grade)

10% Code review

Code quality	/8
Code is well organized : modular code design (different files for data processing, model definition, model training, model inference, …), no/minimal code duplication Evaluation : -0.1 per error	/1
Code respects the PEP8 standard Evaluation : -0.2 per error	/1
Code is easy to read : complex code blocks are accompanied by comments, comments are relevant (see article), functions and variables have meaningful names that reflect their function. Evaluation : -0.1 per error	/1
Proper management of input arguments in the training script (see argparse, python fire, configparser) Evaluation : 1/1 for overall adequate input management, 0/1 otherwise	/1
Proper utilization of GitHub (e.g. branching, relevant commits and messages, usage of pull request) Evaluation : 1/1 if git branches and Github pull request are used for most contributions to the repository. 0/1 otherwise.	/1
Executable scripts have a “main” function (see article) Evaluation : the points given will be equal to the ratio of executable scripts with main function over the number of executable scripts in the project.	/1
Reproducible experiments (e.g. seed) Evaluation : -1 per error	/2

20% Report evaluation

Introduction (~1 page)	/1
Purpose : introduce the project (present and motivate the problem, summarize your contributions, outline the following sections of the report) Evaluation criteria : ● Task is introduced and described ● Model evaluation metric is explained	 /0.5 /0.5
Data analysis (~1 page)	/3
Purpose : in-depth analysis of the provided data using statistics, visualizations, etc. Evaluation criterion : ● Relevant data attributes are analysed : number of samples, available features, missing/corrupted values, underlying structure, and any other data property relevant to the project.	 /3
Literature review (~1.5 page)	/3
Purpose : identify and summarize any prior work (research articles, datasets, models, …) that are relevant to the project Evaluation criteria : ● Literature review is relevant to the project and demonstrates a reasonable level of effort ● Links are drawn between articles when relevant	 /2 /1
Methodology (~2 pages)	/4
Purpose : describe your experimental pipeline Evaluation criteria : ● Experimental pipeline is described in sufficient amount of detail to allow other people to run equivalent experiments : description of models, external data sources (if needed), data splits, data processing, data augmentation, loss function, learning procedure, hyperparameter tuning, … ● The selected methodology makes sense (coherent with the results of the data analysis and the literature review).	 /2 /2

Results and discussion (~2 pages)	/6
Purpose : present your experimental results and analyse them. What can be learned from your results? What works and what doesn't work? How could the results be improved? Evaluation criteria : ● Results are presented clearly using the appropriate media (tables, figures, etc.). ● Analysis demonstrates a good understanding of the results. The presented conclusions are supported by the results.	 /1 /5
Conclusion (~0.5 page)	/2
Purpose : summary of project state Evaluation criteria : ● This section summarizes the work that was done and its limitations ● Recommendations are made regarding future work on the project (what would be the next step, if more time was available?)	 /0.5 /1.5
Overall quality of the report	/1
Report format (7-9 pages including references, font size 11, title with team member names, clear sections, flow between sections, figures and tables titled, axes titled, etc.)	/1

10% Model performance evaluation on blind test set

- If the best model provided by a team crashes or provides results that are statistically worse than those of the baseline model provided by the TAs, the team gets 0%.
- Otherwise, if the best model provided by a team is statistically equivalent to the baseline model, the team gets 6%.
- Otherwise, if the best model provided by a team is statistically better than the baseline model:
    - The team gets 10% if the model is the best performing one or is statistically equivalent to the best performing model provided by another team.
    - Otherwise, the team gets 8%.

Do not attempt to obtain the blind test set in any way (abusing a security loophole, attempting to recover the original data from the internet, etc.). We reserve the right to reject a model if we have valid reasons to believe that any part of the test set was used to build/refine the model.

10% for project presentation (15 min presentation + 5 min questions)

Content of the presentation	/4
Description of the project	/1
Description of the solutions adopted	/1
Presentation of the achievements	/1
Synthesis of findings, obstacles, and recommendations	/1
Format of the presentation	/4
The presentation is clear and structured	/1
The presentation is engaging and speakers have good delivery (looking at the audience, adequate voice level and speaking rate, speaker enthusiasm, ...)	/2
Respect of time	/1
Questions period	/2
The answers to the questions are precise and clear	/2