Course schedule

Lecture	Date	Content	Quiz done before the class	HW	Project
1	Friday Feb 7	ML concepts	-		
2	Friday Feb 14	Regression	1	HW1 out	Project discussion
3	Friday Feb 21	Classification	2		Project preferences
4	Friday Feb 28	Data preprocessing ML pipeline	3		Projects assigned
5	Friday Mar 7	Decision Trees Ensemble learning	4	HW1 due HW2 out	
6	Friday Mar 14	Unsupervised Learning	5		

2-week break from the lectures

Course schedule

Lecture	Date	Content	Quiz	HW	Projects
7	Friday Apr 4	Ethics & model interpretability	6	HW2 due HW3 out	
8	Friday Apr 11	Neural networks	7		
		2-week break from the lectures			
9	Friday May 2	Text mining	8	HW3 due	
10	Friday May 9	Intro to Large Language Models	9		
11	Monday May 12 (makeup from 25.4.)	Applications of LLM	-		Final projects video due
12	Friday May 16	Final project discussion MANDATORY ATTENDANCE	-		Final project code due

Project deadlines

- **Monday May 12th, 12:30:** Project Video due
- **Friday May 16th, 12:30**: Final project notebook due & video vote & peer evaluation (optional)

Attendance mandatory for **Friday May 16th**

Instructions

At the end of this presentation there are instructions for how to:

- Submit your video
- Vote for your favorite video (winner gets extra point(s) for the project grade)
- Submit your final notebook

Project notes

Make sure that you:

- follow the instructions in the project instructions and project template
- include: problem definition, EDA, modelling, NN, evaluation, interpretability, results discussion…
- justify your decisions: choice of tuning metric, choice of best model…
- submit a single notebook per group named: *GROUP_X_Project.ipynb,* the notebook is executable & self contained
- submit peer evaluation if you feel the workload was unevenly distributed (two options: extra points for the MVTM, or lowering the grades of some team member(s))
- watch the videos of the other groups before the class and **vote for the favorite video (best video gets extra points)**
- prepare for a 3 min per group Q&A session (group members will be selected randomly to answer questions, all team member should be present, all team members should be familiar with the whole project)

Video notes

- The video should be 4-5 mins long
- The video will be graded, but it is only a smaller part of the overall project grade
- It will be graded based on how:
  - informative it is about the project (does it explain the most important things about the problem, approach and results)
  - interesting/creative it is as a project pitch
  - extra point goes to the best video vote
- The results presented in the video can be different from the results in the final project submission, so you can still change the project if needed after the video submission

Team member evaluation: increasing the grade of the Most Valuable Team Member

- Each team member will have the possibility to nominate a team member if they feel that they have contributed more than the other team members.
- If a team member gets most of the team votes for being the MVTM, their grade may be increased.
- **Nominating MVTM is not mandatory, should be done for the cases when the workload was not equal.**

Team member evaluation: lowering the grade of the non-contributing team member

- Each team member will have the possibility to evaluate other team members on their **overall participation and contribution to the project**. So, if they feel they all participated equally, they can grade each other with 20. This will be the default value assumed, if the team members do not grade each other.
- To get an individual project grade: project grade will be weighted with the average of the grade given by the other team members. Example:
  - The project gets a grade 18.
  - Two team members, A and B, give the third member C, the following grades: 19 (graded by student A) and 18 (graded by student B), as student C did not participate equally in the project. Hence the student C gets the following grade:
  - 18 (*project grade*) x ((18+19)/2)/20 (*member contribution*) =18 x 18.5/20= **16.65**
  - **Note that if you grade a fellow teammate with 18 that does not mean that they should get an 18, only that whatever the final project grade is, it should be weighted with 18/20.**
- **Hence, any grade less than 20 will lead to a decreased grade of a fellow team member**. Please be mindful of this when you evaluate each other!

Video Submission

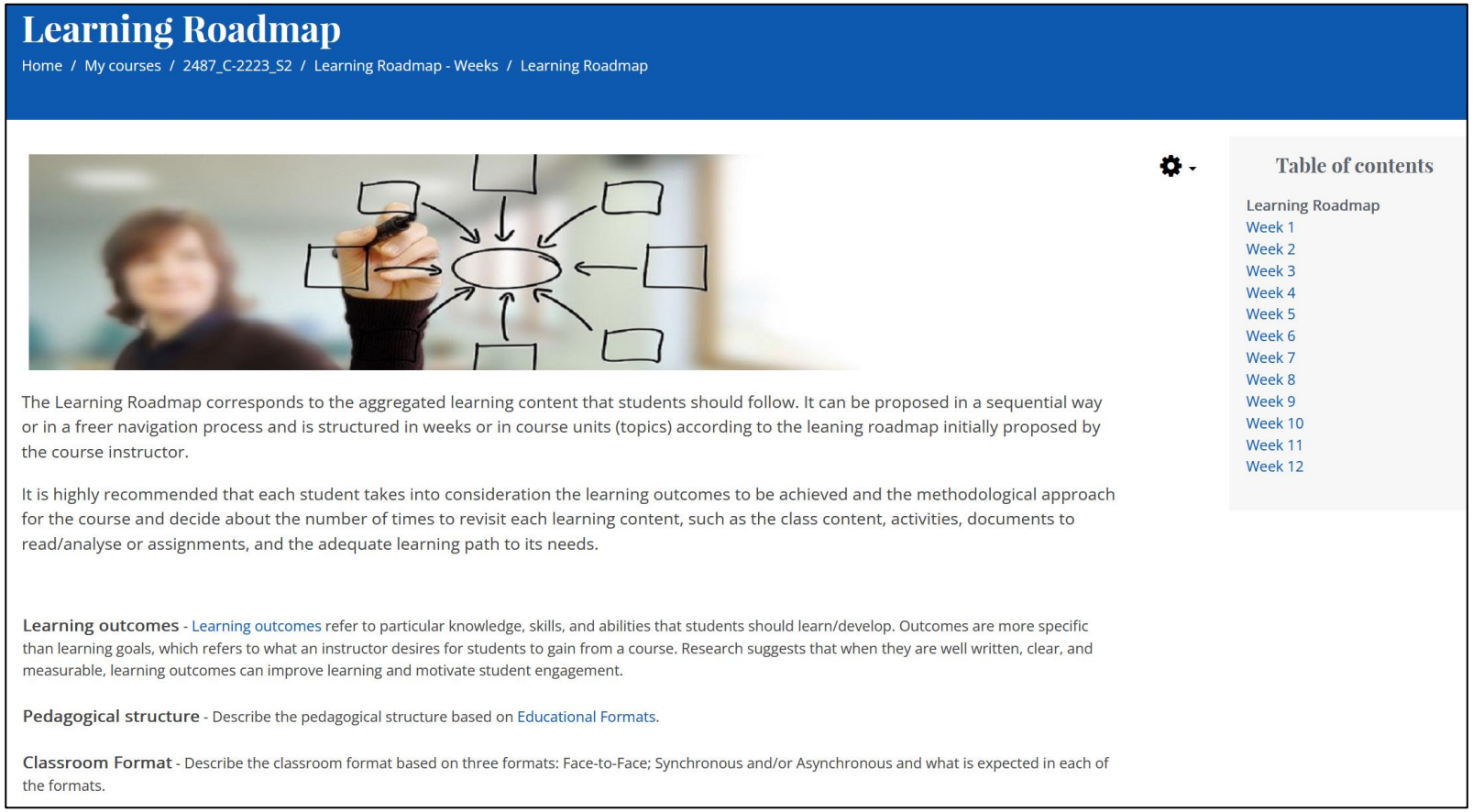
STEP 1

- In the Moodle page, select “Learning Roadmap”.

Video Submission

STEP 2

- In “Learning Roadmap”, select “Week 11”.



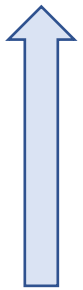
Video Submission

STEP 3

- In the “Activities” section, select “Video Submission”

Activities

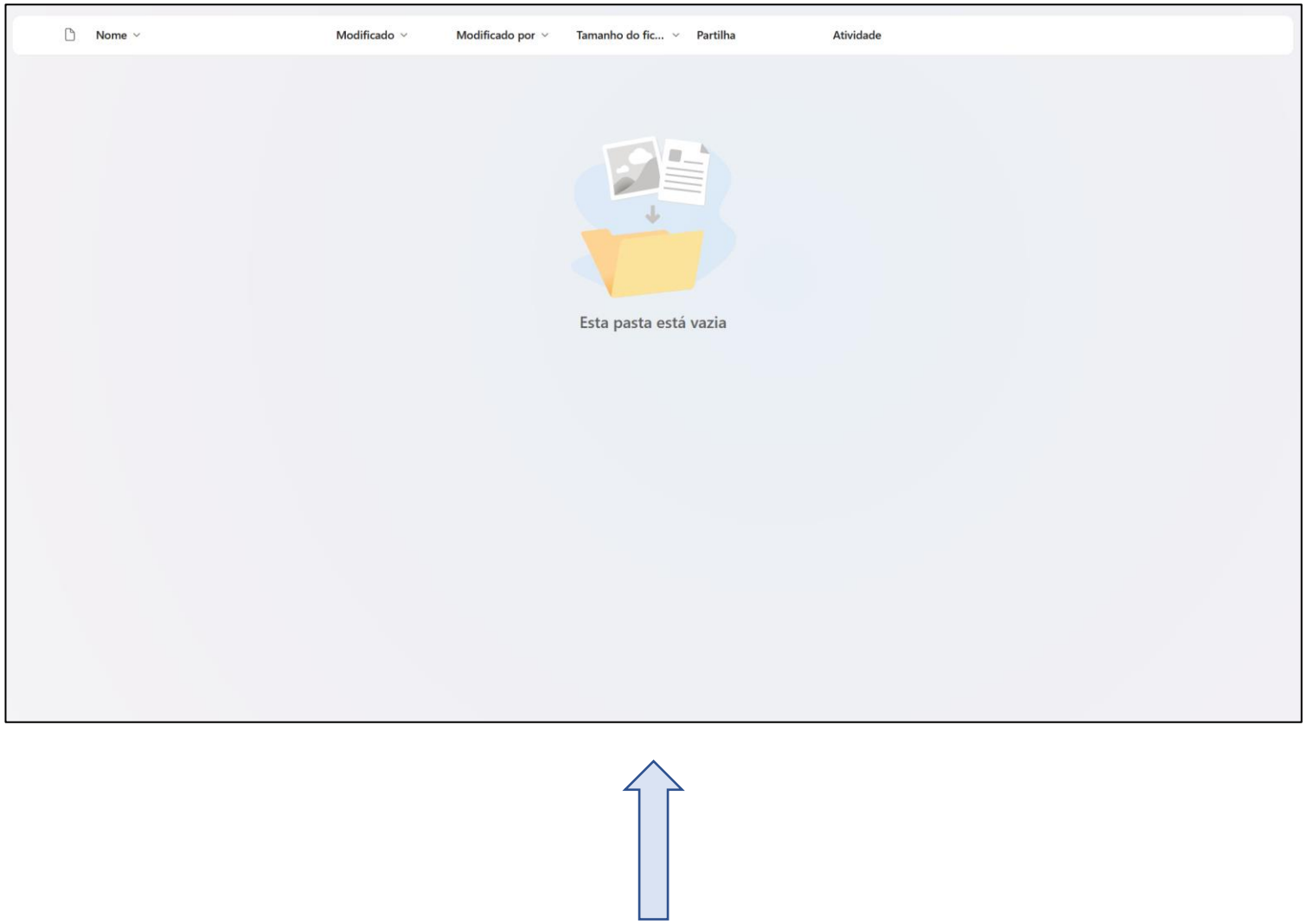
	Access	Deadline
Video submission	*Submit your 5 minute video here* as "Group <group number> - <dataset name>" (e.g., Group 1 - House Price).	*12th May 12h30*



Video Submission

STEP 4

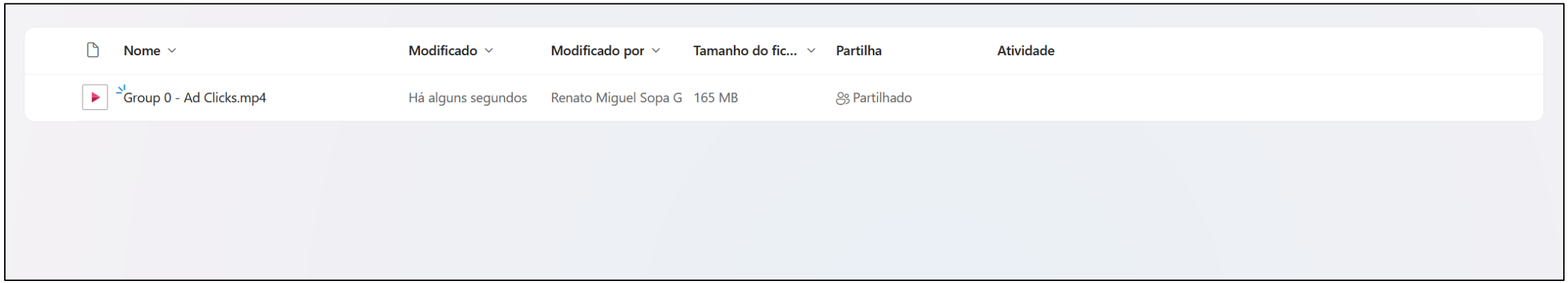
- In Sharepoint, drag your **5-minute** video onto the page to upload it.



Video Submission

STEP 5

- After a while, the video will be displayed on the channel. To view it, click on it (recall that the deadline to submit the video is on the 12th of May, 12h30).



Best Video Vote

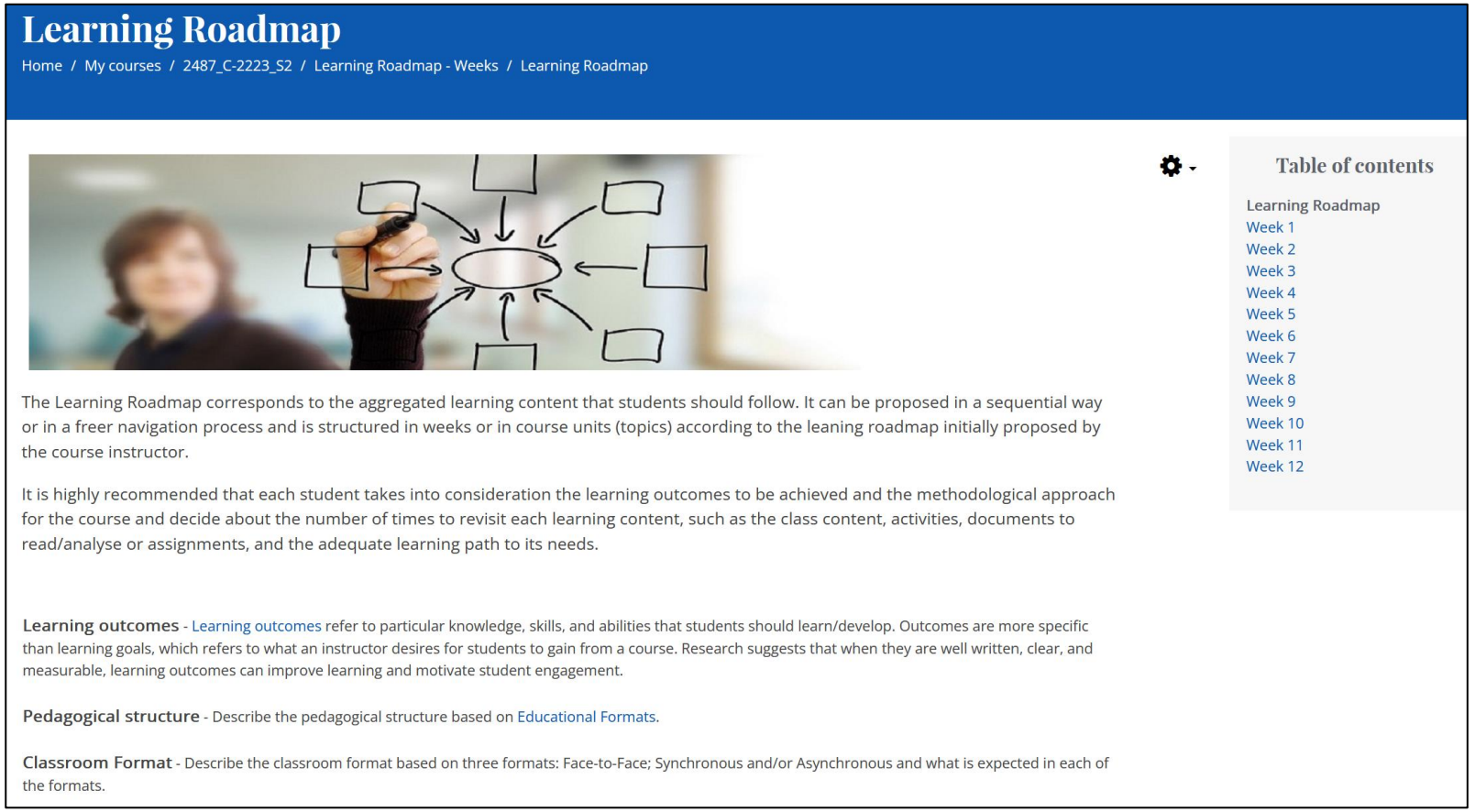
STEP 1

- In the Moodle page, select “Learning Roadmap”.

Best Video Vote

STEP 2

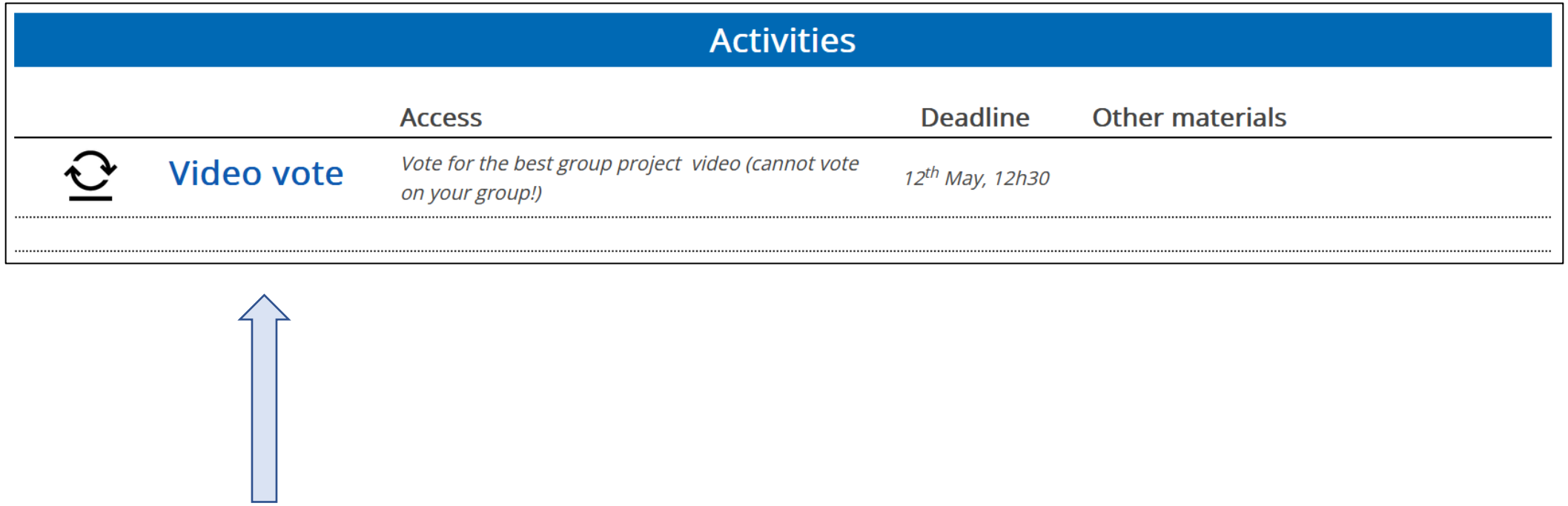
- In “Learning Roadmap”, select “Week 12”.



Best Video Vote

STEP 3

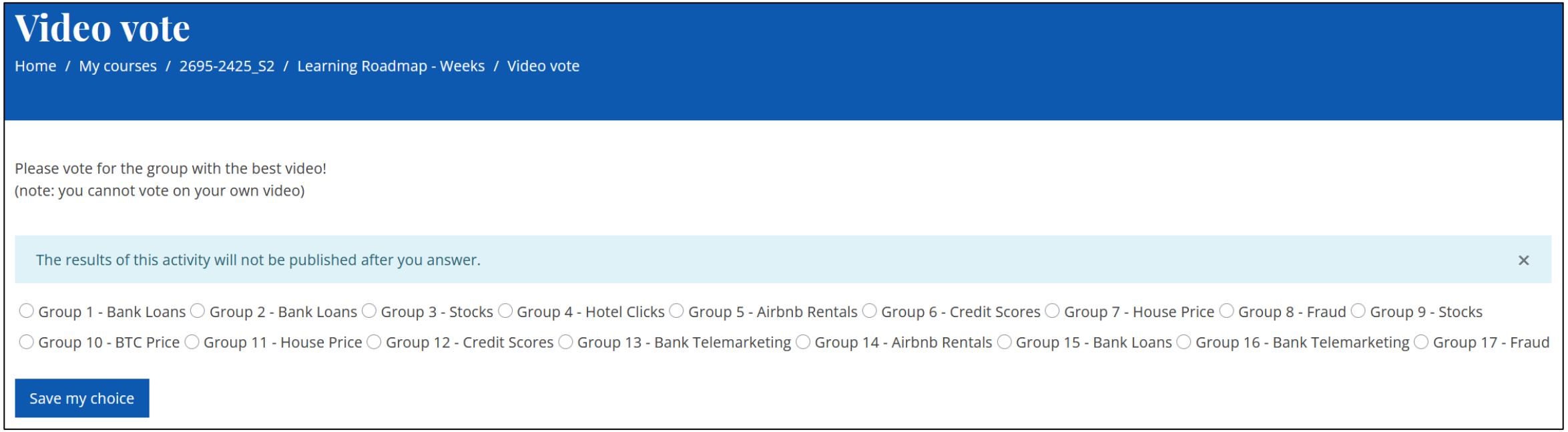
- In the “Activities” section, select “Video Vote”.



Best Video Vote

STEP 4

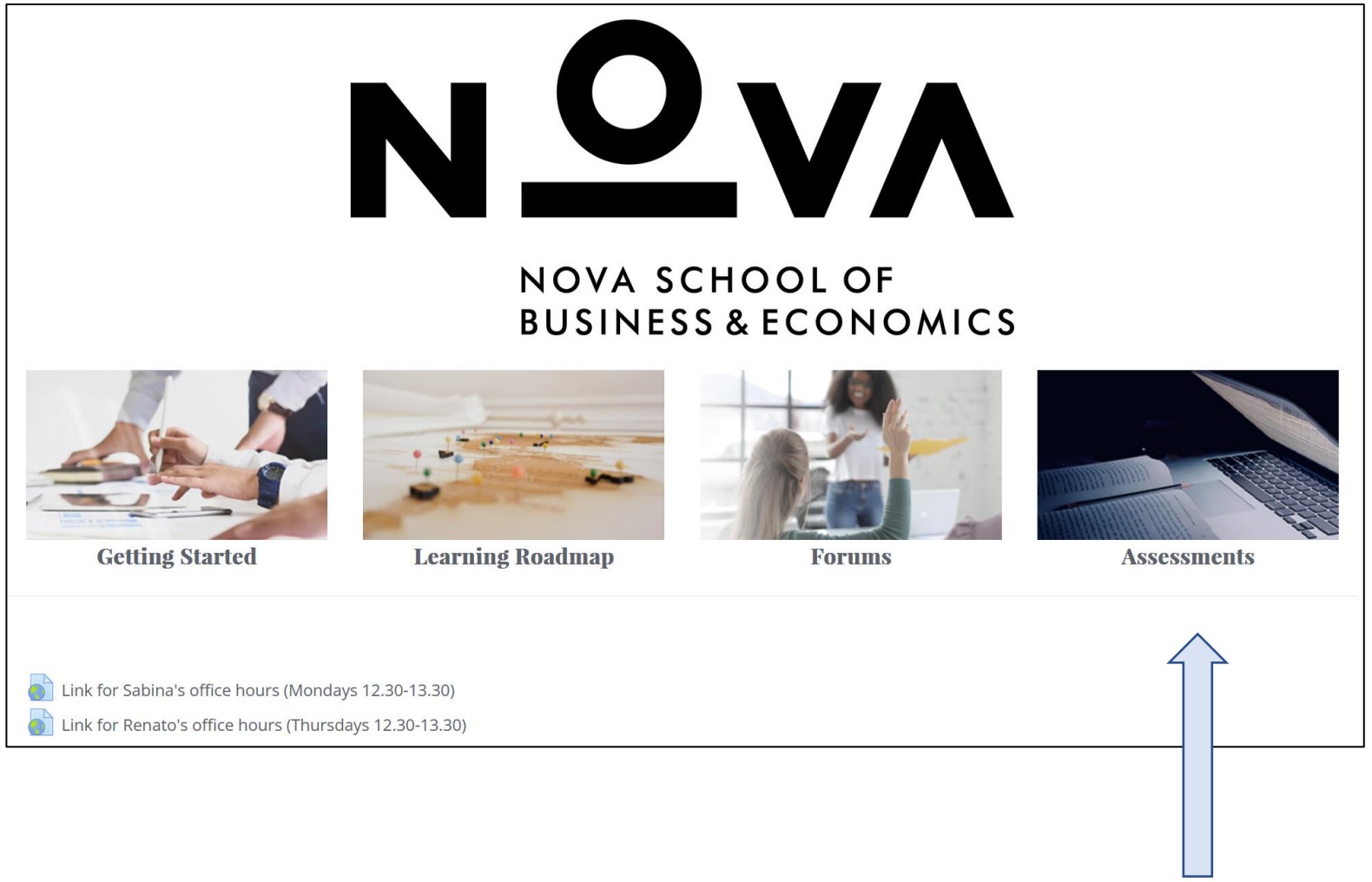
- In “Video vote”, select the group with your most preferred video;
- Save your choice;
- Edit before the deadline (16th of May, 12h30), if warranted.



Project Submission

STEP 1

- In the Moodle page, select “Assessments”.



Project Submission

STEP 2

- In “Assessment details”, select “Group project – final – final submission”.

Summative assessment is a vital component to evaluate how much a student has learned throughout a course. Taking that in mind, please be prepared by assisting classrooms, developing the proposed activities and be active during all the other components of the course strategy. And do not forget to interact with the other students sharing knowledge and learning all together.

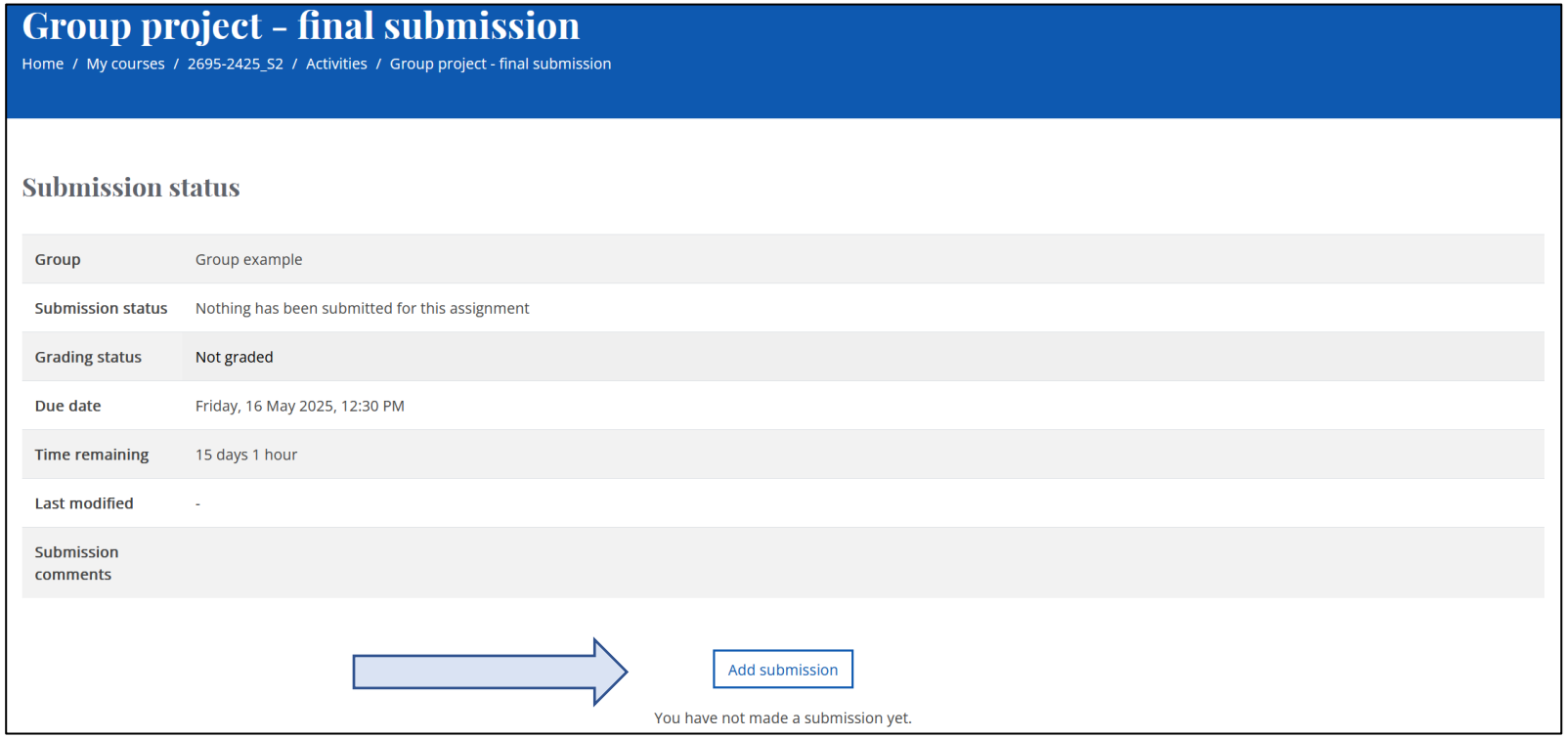
Assessment details

Evaluation moment	Available date	Due date
Homework 1	February 15th, 2025	March 7th, 2025
Homework 2	March 8th, 2025	April 4th, 2025
Homework 3	April 5th, 2025	May 2nd, 2025
Group project - final submission	-	May 16th, 2025
Group project - final - peer evaluation	-	May 16th, 2025
Theoretical exam	-	-
Coding Exam		

Project Submission

STEP 3

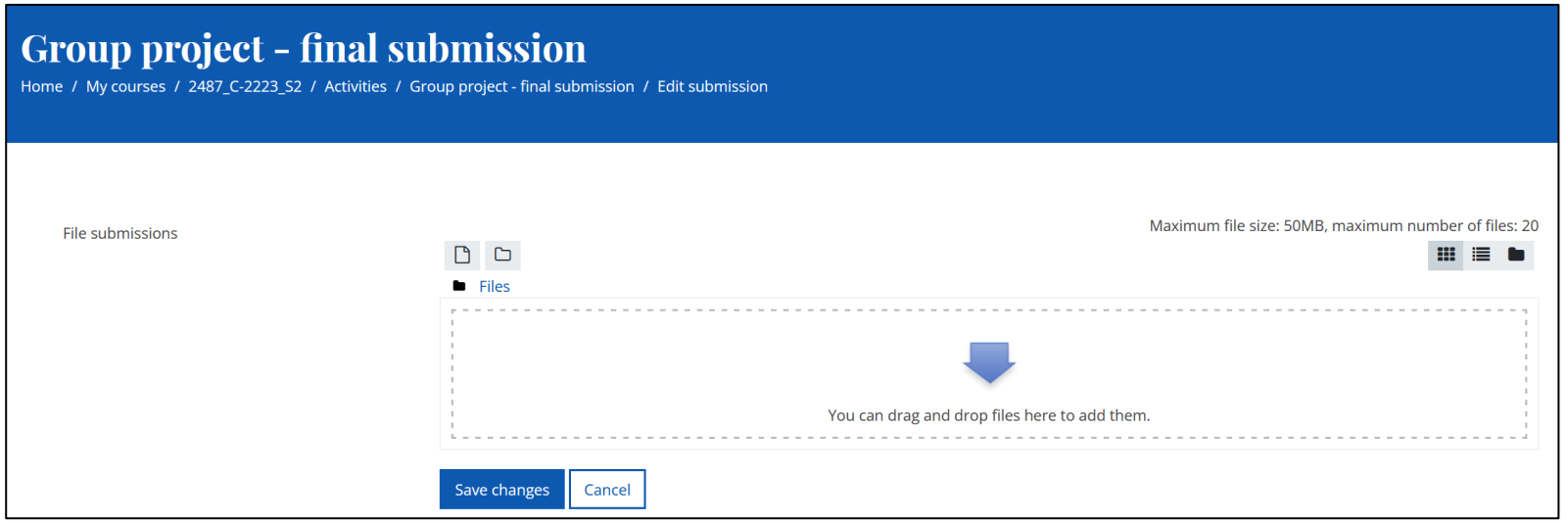
- In the “Group project – final submission” page, select “Add submission” (should adjustments be necessary, “Edit submission” is available until the deadline).



Project Submission

STEP 4

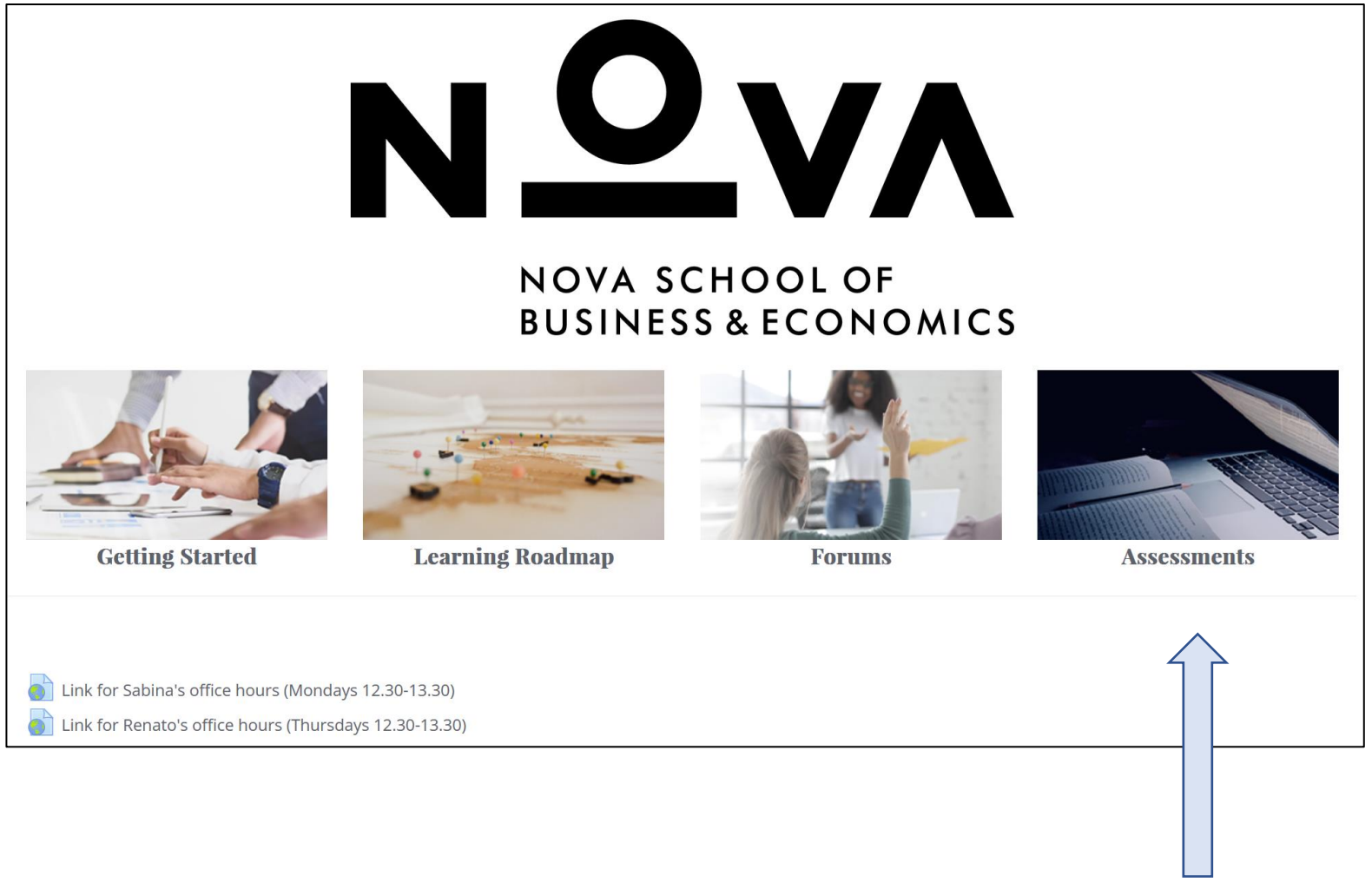
- In “Group project – final submission”, upload your project and the associated dataset;
- Save changes;
- Deadline is on the 16th of May, 12h30.



Peer Assessment

STEP 1

- In the Moodle page, select “Assessments”.



Peer Assessment

STEP 2

- In “Assessment details”, select “Group project – final – peer evaluation”.

Summative assessment is a vital component to evaluate how much a student has learned throughout a course. Taking that in mind, please be prepared by assisting classrooms, developing the proposed activities and be active during all the other components of the course strategy. And do not forget to interact with the other students sharing knowledge and learning all together.

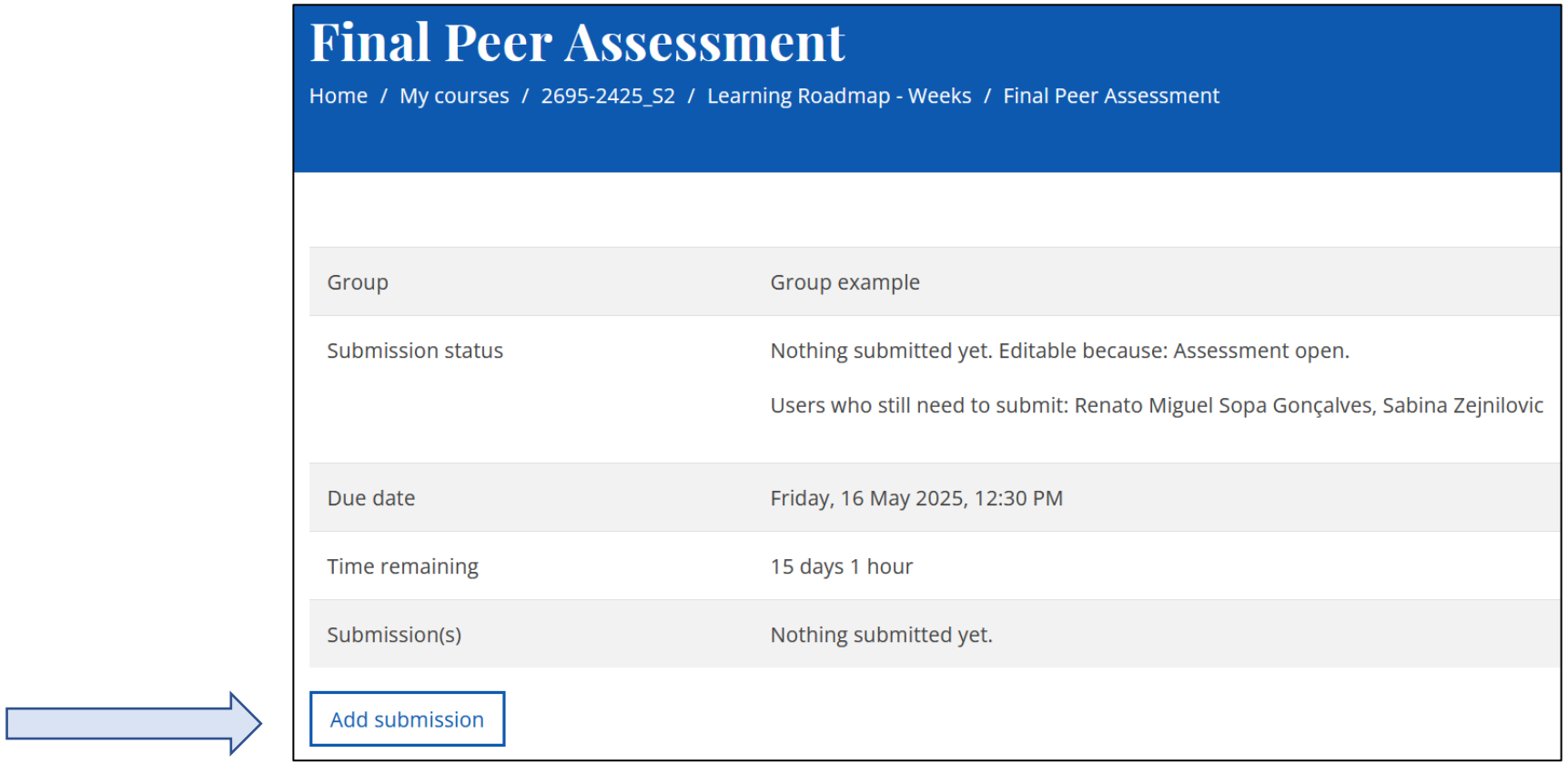
Assessment details

Evaluation moment	Available date	Due date
Homework 1	February 15th, 2025	March 7th, 2025
Homework 2	March 8th, 2025	April 4th, 2025
Homework 3	April 5th, 2025	May 2nd, 2025
Group project - final submission	-	May 16th, 2025
Group project - final - peer evaluation	-	May 16th, 2025
Theoretical exam	-	-
Coding Exam		

Peer Assessment

STEP 3

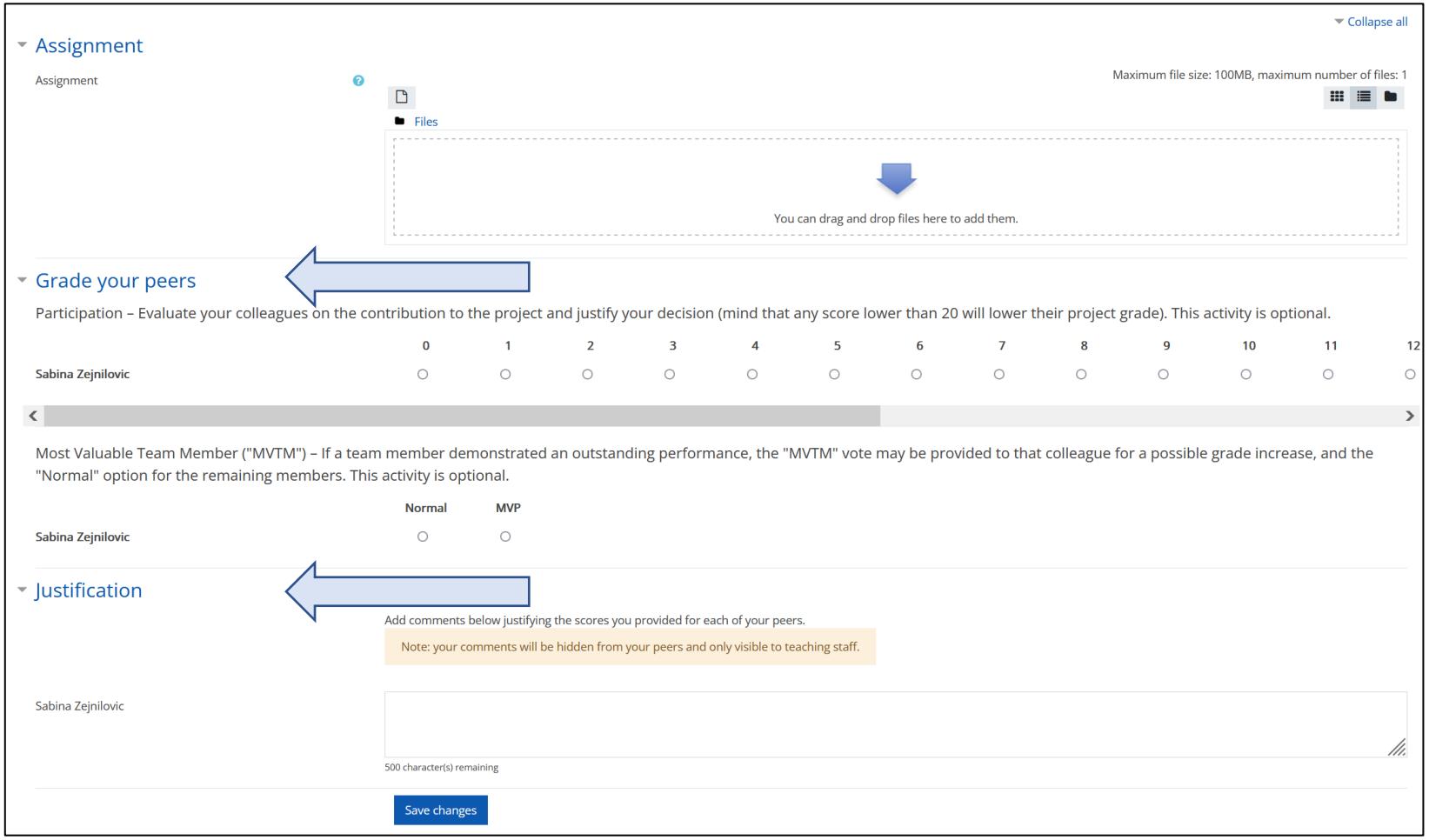
- In the “Final Peer Assessment” page, select “Add submission” (should adjustments be necessary, “Edit submission” is available until the deadline).



Peer Assessment

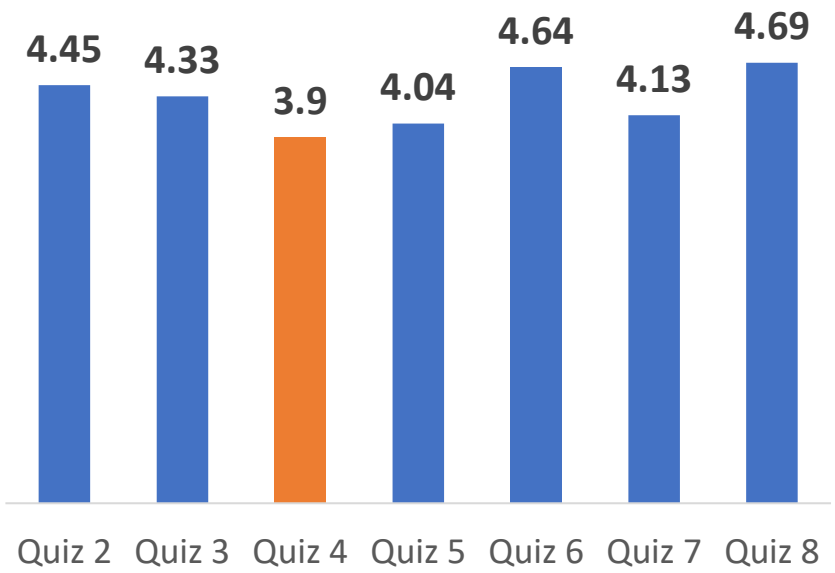
STEP 4

- Ignore the “Assignment” category;
- For each of your team members, select a grade from 0 to 20;
- Note that any grade less than 20 will lead to a decreased grade of the fellow team member, as outlined in week 2;
- If a specific team member performed exceptionally, the MVTM vote may be awarded;
- Justify the grades and MVTM vote provided to your team members;
- Save changes;
- Edit before the deadline (16th of May, 12h30), if warranted.



Quizzes 2-8 Summary

Quiz Averages

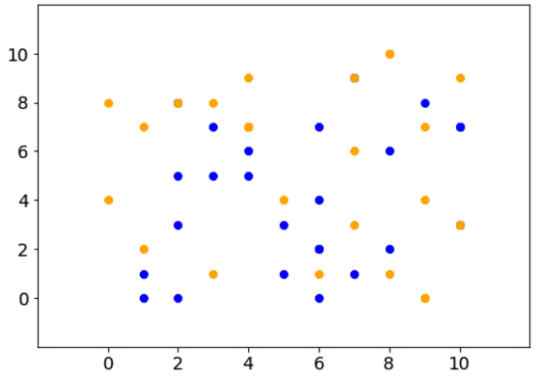


Average Grade Overall: **4.31**

Most Difficult Quiz 4 Questions

Scaling Method (**74% correct**)

none: not zero mean, not 0-1



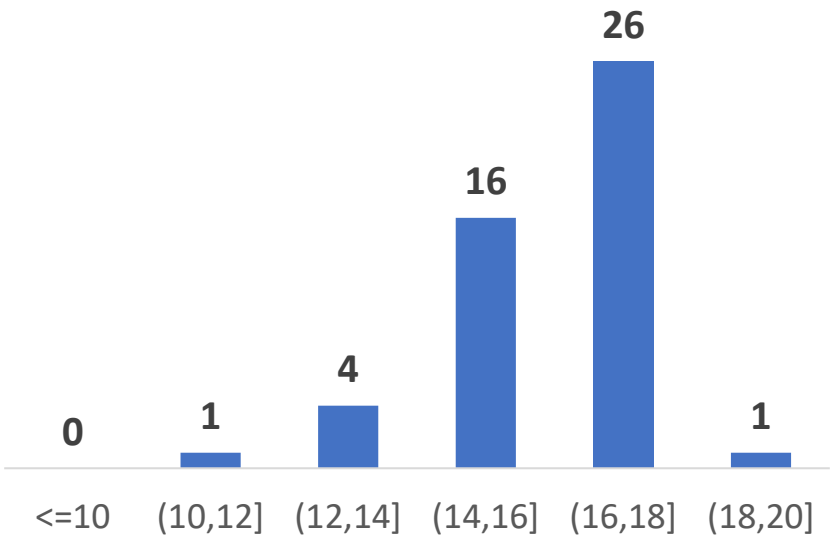
Encoding Categorical Features (**34% correct**)

Feature name	Type	Values it can take
Number of visits	Numerical	Integers from 1,2,3…
Favorite color	Categorical	Red, Green, Blue, Yellow
Income	Numerical	Any positive float value
Age group	Categorical	Child, Teenager, Youth, Middle Age, Old

1 (numerical)+ 3 (colour=nominal, non-collinearity)+1+ 1 (age=ordinal)

Homework 1 Summary

HW1 Results



Minimum: **11.58**

Maximum: **18.42**

Average: **16.02**

Most Difficult Question

2.6. Train the logistic regression with solver='liblinear' with regularization strength equal to 0.01 with lasso regularization and that stops converging after 700 iterations (2pt)

```
#Standardization could be performed but it was not required.

from sklearn.linear_model import LogisticRegression

log_reg = LogisticRegression(solver = 'liblinear', C = (1/0.01), penalty = 'l1', max_iter = 700)
log_reg.fit(X_train, y_train)
```

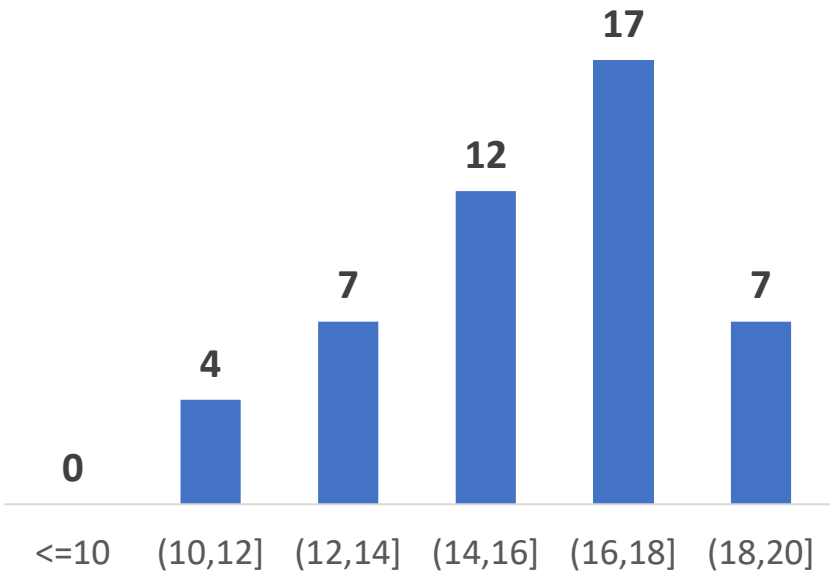
Question 2.6 (**18% correct**)

Recall that Logistic Regression requires the

inverse of the regularization

Homework 2 Summary

HW2 Results



Minimum: **10.38**

Maximum: **18.85**

Average: **15.64**

Most Difficult Question

2.4. Using pipeline, do the following (not necessarily in this order, but choose the right order of steps in the pipeline)

- Use SMOTENC for balancing
- Normalize the numerical variables
- Perform one-hot encoding for all the nominal variables by keeping all the dummy variables
- Use mean imputing for numerical variables and mode for the nominal
- Use randomized search with random state 42, and 25 combinations of parameters to select the best classifier: try Random Forest and CatBoost
- Use ROC Area under the curve as the criterion for model selection and tuning
- The cross validation needs 5 splits that shuffles the data the same number of times and keeps the same proportion of target classes, the test size of 0.3 and a random_state = 42

Tip: after using the column transformer, the column names are replaced with indexes

Question 2.4 (**58% correct**)

Homework 2 Summary

SMOTENC

- Balances an imbalanced dataset by synthetically generating new samples of the minority class.
  - numerical features: interpolates between nearby minority-class examples
  - categorical features: uses a mode (most frequent category) of neighbors, so OHE should not be done before SMOTENC
  - uses distance calculation: scaling should be done before SMOTENC