

P2 – ML Applications in Digital Cultural Heritage (12009 & 12017) – Examination Summary and Grading

Dr. Anastasia Eleftheriadou & Dr. Evan Eames

November 24, 2025

Students enrolled in both the Seminar and Hands-On (Übung) component of the P2 module are examined together with a common assignment. The goal is for students to attempt to showcase an archaeological / cultural heritage case study by making a short review of the existing literature and doing a hands-on assignment working with messy real-world datasets by employing data cleaning and ML ideation.

Be sure to read the tips on the final page of this document!

Examination Procedure

1. Students choose a dataset from a list <https://tinyurl.com/5ddxezjk>. There is one dataset per student – that is, no two students can work on the same dataset. For your convenience, when you’ve put your name down for a dataset (at the previous link) you can find the actual data files here: <https://tinyurl.com/mtkxj67v>
2. Students conduct a literature review around the dataset. The goal here is to get ideas for how international research teams have used similar data for interesting research purposes (note: it is not necessary that the articles actually employ ML techniques).
3. Students clean and manipulate the dataset. This should be done in a notebook file with each step nicely explained in markdown. The goal is to bring the dataset to a state conducive to further exploration (plotting, ML, etc.) which could be useful for Archaeology.

4. Students write an accompanying paper (3000 - 5000 words). This paper must comprise the following sections:
 - (a) An introduction to the state-of-the-art relating to the topic of the dataset you've chosen (pottery, lithics, osteoarchaeology, zooarchaeology, archaeobotany, dating, etc.)
 - (b) A summary of the specific dataset you've chosen, as well as a discussion of how it has been cleaned and prepared. This should include a section on difficulties faced in cleaning
 - (c) A discussion of imagined subsequent applications which could now be attempted with your dataset, and how you would attempt them. These should be based on what similar teams have attempted based on your literature review. You should discuss expected challenges and how you would deal with them.
 - (d) **Note:** Any students who successfully implement ML applications will receive bonus points – more points for more relevant and ambitious applications.
 - (e) Finally, a conclusion summarizing the entirety of the project.
5. Students must submit their papers and notebook files together by 23:59 on February 6th. 10% will be deducted for each day beyond this deadline.
6. In the final 2 weeks of the semester students will be invited to a 30 minute “mini-defense”. This will consist of Anastasia, Evan, and each student. The structure is:
 - (a) 5 minutes of presentation of the state-of-the-art
 - (b) 10-15 minutes of overview of what has been done for the project (powerpoints can be included, but are not required).
 - (c) 10-15 minutes of questions
7. Additionally, in the final classes each student will do a quick 5 minute lightning summary of their work (not graded).

Grading

| | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 |
|-------------------------------|--|--|---|--|---|
| Paper (20%) | Exceptionally well researched, written, and relevant to the dataset | Well researched, written, and relevant to the dataset | Decently researched, written, and relevant to the dataset | Poorly researched, written, and relevant to the dataset | No effort has been made |
| Notebook (20%) | Beautifully presented with clean markdown, figures as needed, and fully commented code | Well presented | Decently presented, but some difficulty in understanding | Poorly presented and messy. Difficult to understand | No effort has been made: It looks like sh*t |
| Mini-Defense (60%) | Interesting and engaging presentation demonstrating the topic has been thoroughly analyzed. Questions are answered thoroughly with in-depth reference to course material | Good presentation, with answers sometimes reflecting an understanding of course material | Decent presentation, with satisfactory answers to questions | Poor presentation, showing a lack of understanding of the dataset and topic, as well as difficulty answering questions | Complete lack of understanding. |

Tips

- Be sure to have plenty of citations in your paper. Be sure it follows the format mentioned in the above section.
- Be sure that your notebook is nicely presented. This means markdown explaining sections, nicely commented code, and figures where appropriate. You can use the course notebooks on moodle as inspiration
- Be careful if your .csv file contains items separated by semi-colons “;” as opposed to commas – in this case you should use the `sep=”;` argument with `pd.read_csv`
- It’s very possible there will be lots of new things in the dataset we haven’t seen in the class yet (ie. multiple sheets, multiple headers, different formatting, etc.). This is expected. You need to go online and try to figure out what to do.
- In the presentation we will be testing your knowledge across the entire course. It’s worth going through the papers from the seminar course (at least those relevant to your topic) and also review the notebooks from the hands-on course (including the exercises)
- Feel free to come by our office for help if you need any guidance or advice.