Data Science Project (PX55PA/PB)

Description

The data science project is a very important component of your training. It will provide you with experience in conducting an independent research project. It will constitute an opportunity to apply the knowledge and understanding that you have gained and the tools that you have learnt during the courses in semesters 1 and 2 directly in a research setting. Moreover, it will enable you to acquire a range of important planning, analytical, oral and written presentation skills.

Overall course assessment

There are 3 main components to assess the Data Science project:

- Oral presentation (mandatory but it does not contribute to final mark): 20 min oral group presentation explaining the motivation of the project, main objectives and workplan.
- Thesis (65% of overall mark): written thesis to be submitted by the 29th of March (full time)/21st June (part-time), see details below.
- Viva (oral exam, 35% of overall mark): vivas will take place in the week of the 15th of May (subject to changes)/beginning-mid July for part-time students.

Structure of thesis

The thesis will consist of 2 parts:

1. Written report (max 4,000 words, excluding references, abstract and captions of figures and tables; please provide final word count): the motivation, background, methodology, main results and discussion need to be presented in this report in the format of a scientific paper (in word or pdf format). The report needs to include the following sections:

    - Title of the project, student's name and ID number
    - Abstract (about 300 words, not included in word count): it should describe the main objectives of the project, how the study was done (without methodological detail) and summarise the main results and their significance.
    - Introduction: provide background to put the research conducted in your project into context and to allow readers outside the field to understand the aims and significance of it. Give an overview of what has been done in the literature on this topic.
    - Data: describe the data used in your analysis.
    - Methods: technical details about the methods used (definition of main concepts, main equations, parameters) should be given so that a researcher in the field can reproduce your results.
    - Main results: this section should describe the main results of your research project, the interpretation and conclusions that can be drawn. You should support this section with the main figures and/or tables illustrating your results. What actionable insights can be gained from your analysis? If you include further figures and or tables in the

Supplementary Material section (please see point 2 below) to support your main results, you should make refer to it in the main text and make clear where those results can be found (e.g. see Table S2 from Supplementary Material).

- Discussion: this section should remind the reader about your aims, main results and their implications. If applicable, you should compare your results to previously obtained results in the area and you can suggest next steps of future research following up from your results.
- References: you should cite all sources that you have used for your research (published papers, pre-prints, webpages, codes, etc), as well as data sources used. Please note that a reference should be given at every point in your report where you cite published information. There are many schemes for referencing material. The important thing is that you are consistent. There are many resources provided by the university to help you with referencing in academic writing, please check this link (https://www.abdn.ac.uk/students/academic-life/study-resources-3379.php#panel3953). References do not count towards the word count.

Given the word limit of the written report, you will need to write in a concise way, highlighting the main points of the research that you have carried out during your project and selecting the key figures/tables that support your most important results. The rest of the figures relevant to the project can be included in the Supplementary Material (please see next point). Tables and Figures need to be numbered, so that they can be referred to in the text, and they also need to include a caption explaining their content.

The university provides many resources to help you developing your academic writing skills. Please check the following link, under the tab Academic Skills: Student Learning Services, Academic Writing: Come to the Writing Room! (https://www.abdn.ac.uk/students/academic-life/study-resources-3379.php#panel3953). There are academic writing sessions that you can book through the online course booking system.

2. Supplementary Material (no length limit): the supplementary material should include the main codes that you have written to carry out your research. You should comment the codes in detail, so that the different steps that you have coded are clear to another person not familiar with the code. Moreover, the code needs to be in a runnable form. You can also include supplementary figures and tables that support the results that you have presented in the main written report. In that case, it is very important that you refer to the supplementary material in the main written report. You can also add more technical details, such as information about the database and further computational resources that you have used. For example, if you have built an SQL or Mongo database, you should explain your design decisions. The supplementary material can also contain analyses that did not directly lead to insight. The supplementary material can be submitted as one or several files.

Your supervisor will also help you to develop your scientific writing skills by giving you feedback on the first draft of your thesis. For this to properly work, however, it is crucial that you send a first draft of your thesis well before the thesis submission deadline. This is to make sure that your supervisor has time to read your draft and give you feedback. Please discuss with your supervisor an appropriate timeline for this to occur.

Plagiarism

Copying or reporting the work of others without attributing the source is plagiarism and, if detected, may lead to failure. Please see notes above about how to write references in your written report. Please also check the following university link, which explains how to avoid plagiarism (https://www.abdn.ac.uk/students/academic-life/study-resources-3379.php#panel3954)

Format and timeline of viva (oral exam)

The viva or oral exam will take about 20 minutes. Your supervisor and an internal examiner will be present (the external examiner might be present too). The internal examiner will lead the oral examination. The role of the supervisor is to provide clarification (only if needed).

The viva is a session where the internal examiner will ask you questions about your thesis. The purpose of the viva is to test your understanding of the topic of your thesis, the methods that you have used to analyse the data, how you interpret your main results and which conclusions can be drawn from them. In preparation for your viva, you should read again your work with a critical eye and think about areas of the work that might prompt questions. You should make sure that you understand all aspects that you have written in your thesis, and discuss the viva with your supervisor before the date of the oral exam.

Sometimes students tend to view some of the questions as a challenge, so that they fail to notice how simple the question actually is. Try to listen calmly, and do not rush your answer. If you do not understand the question, please ask the examiner for clarification. Be prepared to answer technical questions, as not all examiners will be familiar with the methods and tools that you have applied in your project. In the case that you do not know the answer to a question, it is not the end of the world. Just try to keep focused on the rest of the questions.

Further details about how the oral exams will be conducted, will be made available closer to the date.

Assessment criteria

Oral presentation:

Presentation: prepare clear slides, legible, with good balance of graphics and text. It is also important that you plan the structure of your talk, so that it is logical (audience can understand it even if they are not familiar with the topic). Keep the talk interesting, motivate the topic of your research project and highlight open questions. Keep the information presented relevant to the chosen topic, and make sure that you have good timekeeping.

Content: make sure that you have the right amount of introduction to the topic for peers; that the explanations are clear and correct; that the data science concepts are clear and at the right level, and that you present a clear and logical plan for the project.

Thesis:
The criteria that will be use to assess the written thesis will include:

- Presentation/structure
- Literature (analysis and coverage)
- Critical analysis
- Understanding of material
- Ability to communicate
- Quality of work
- References (correctly presented)
- Word limit
- Supplementary material and comments provided on the submitted codes

Oral exam:
The criteria that will be used to assess the oral exam will include:

- Understanding of material, depth of knowledge.
- Ability to communicate ideas
- Breadth of knowledge

The thesis and oral examination will be assessed by both the examiner and supervisor, who together will agree on a mark for each of those components. Feedback in the form of written comments will be sent to the students.

Coordinators: Marco Thiel and Mamen Romano