CSC8634: Cloud Computing with Project - Extended Technical Project

Semester 1, 2020

Aims

The aims of undertaking this Extended Technical Project as part of **CSC8634 Cloud Computing with Project** are to give you:

- an opportunity to carry out research to deepen your knowledge and develop your skills regarding an area in which you have a special interest
- an opportunity to develop your skills regarding report writing
- experience of managing a data science project.

Project Overview

For this assignment you will work independently [1] on a substantial project relating to Cloud Computing.

You are strongly encouraged to make a project selection which most closely aligns with your interests, passions, any particular areas of technical expertise you would most like to practice and develop further, and your future career aspirations.

Each Extended Technical Project will primarily be assessed on the rigour with which the problem is tackled, and your personal and professional development associated with undertaking the project. This should give you permission to pursue ambitious projects without worrying about *'hedging your bets'* with datasets you perceive to be *'safer'* or *'easier'*. Should you wish to discuss your options before embarking on the project, please do not hesitate to contact Matt [2].

[1] As usual, while your submissions need to be solely your own work, I strongly encourage you to discuss your coursework with your colleagues, collaborate around ideation and reach out to me at any time you would like advice.

[2] `matthew.forshaw@ncl.ac.uk`

Selecting a dataset/project

Details on projects and datasets selection are presented in a dedicated Github repository [3]. As well as providing details of the projects and access to the datasets, this repository will exist as a *'living document'* and Frequently Asked Questions (FAQ) areas and additional resources will be posted periodically.

This is a private Github repository, so you will receive a `404` error if *a)* you have not been granted access to the repository and/or *b)* you are not logged into Github when you access the repository.

[3] `https://github.com/NewcastleDataScience/StudentProjects202021`

Deliverables

Part 1. Written Report

The main written report for the Extended Technical Project should describe the work you have undertaken, and address the following questions:

- What is the need for the project?
- Justify your choice of response (i.e. the nature of, and your plan for, your project). To give strength to your argument, you should reference to practice elsewhere (e.g. in academic literature, or industry practices).
- What, concisely, did you do?
- How successful has it been? Provide evidence, using appropriate evaluation methodologies, and comment on the strengths/weaknesses of your evidence in answering this question.
- What are the future implications for work in this area? If applicable, which areas of extension work are now possible due to the foundational work you have performed in this project?
- A brief reflection on your personal and professional learning in undertaking this project. Here you can comment on how you found the process, what you learned about the technologies and methodologies you used, which aspects you found most difficult/straightforward, and any conclusions which will inform the way you undertake similar projects in the future.

You should also produce additional documentation detailing the findings from your exploratory analysis. You are encouraged to make use of a literate programming framework, e.g. R Markdown [4] / Jupyter [5], to align analytic code with narrative text. You should submit the source file(s) for the notebook(s) as well as output saved in PDF format. There is no particular limitation on the length of this document, but it should be structured clearly.

[4] http://rmarkdown.rstudio.com/

[5] http://jupyter.org/

Part 2a. Structured Abstract

For many projects you undertake in academia and industry, you will be requested to provide an Executive Summary or Structured Abstract, which is intended to summarise the complete project. These commonly range from 200 words in length to around one page.

For your Extended Technical Project we would like you to produce a Structured Abstract for your project, maximum 200-250 words. Structured Abstracts are a very useful tool to reason about the most important aspects and outcomes of a project, to be able to communicate these effectively to key stakeholders.

To construct a Structured Abstract for your project, write short statements responding to the five headings below. These should be written such that the headings can be removed, and abstract reads as stand-alone text [6].

[6] For the purposes of your submission to CSC8634, please include the headings.

Context: In one to two sentences summarise the background context to your work; clearly state why it is an important problem to study.

Objective: Describe the purpose of your project, clearly state the problem you set out to investigate in the work you have done.

Method: Summarise the technical approach or apparatus you have used or developed in the project.

Results: Describe the results that you have obtained; be as specific as possible and if appropriate quantify the results.

Novelty: Please clearly state the novel contribution of your work in comparison to previous publications in the field. If you have previously published in this area, please explain how this work differs from your previous papers.

Part 2b. Key Images

"A picture paints 1000 words." You should also highlight no more than two *'key'* plots/figures/diagrams which can accompany the Structured Abstract. These should be carefully selected to be standalone (i.e. annotated and labeled such that they do not require an accompanying narrative), and should represent a visually appealing accompaniment to your Structured Abstract text.

In February 2021 we will hold a short training workshop to begin constructing our work-based portfolio ahead of making applications to jobs post-graduation. Your Abstract and Key Images for each of these projects will form the basis of this session.

Best-practice development

Throughout this coursework we are not simply interested in a solution which achieves some stated functionality or a desired outcome. It is also important for you to bring together your experience from earlier Semester 1 modules (e.g. Data Management and Exploratory Data Analysis) and make use of best practices. You will also be assessed on the following:

1. You should make use of **Git** for version control. All development activities should be carried out using Git, i.e. there should be no passing around of source code and materials via other mechanisms such as email. As a rule of thumb; *"if it isn't visible in the version control logs, it didn't happen"*. Please ensure you are using only **private repositories** for your project work. You should also avoid committing raw data to your repository, as Git (particularly at Github's scale) suffers with large binary files.

2. All source code and programs as part of your solution should be **well documented**.

3. You should consider the **reproducibility** of your analysis, making use of appropriate tools such as ProjectTemplate for R.

4. You should produce much of your written documentation using a **'literate programming'** framework such as RMarkdown for R, or Jupyter Notebooks for Python.

Guidelines on Deliverables and Online Submission

You will submit your assignment electronically via NESS [7] by **15:30pm on Friday 22nd January 2021**. Your *'deliverables'* will include...

[7] `http://ness.ncl.ac.uk`

Source code You are expected to submit all source code developed in the coursework. You should also provide a `README.txt` document clearly stating which files relate to which part of the coursework solution. Your `README.txt` file should also provide instructions on running your analyses. These instructions should be sufficient to run the analysis. This should be automated as much as possible, and any non-automated configuration or installation steps should be clearly documented.

Written report Written reports should be submitted in PDF format, and should clearly indicate your name and student number within the document, and also in the file name.

Structured Abstract and Key Images Please also submit your Structured Abstract and Key Images in a separate file to your main report.

It is likely you will be submitting a large number of files, in a nested directory hierarchy. You will find it most convenient to zip these files up prior to submitting them to NESS. Please ensure any zip files contain your student number and module code in the filename.

Oral Examination

After the submission deadline for the Extended Technical Project, we will schedule a structured discussion including a software demonstration and reflection on the key learning objectives of the Extended Technical Project. We refer to this as a *"Zero Weighted"* assessment; this is a mandatory part of the course – and key to your learning – but it does not contribute towards the grade you will receive for the module.

Questions?

If you have any queries about CSC8634 assessment, please contact me on `matthew.forshaw@ncl.ac.uk`.