CSSci

Code Portfolio and Explainer Session

Individual Assignment DE

At the end of each DE workshop from week 3 to week 7 of the semester, you will be tasked with completing a coding assignment related to the method you were shown in class on that week. The task will typically involve an application of the method. Minimal solutions will be shared at the end of each workshop, based on a standard dataset. Your task is to complete the assignment on a **different dataset**, one of your choosing (you can pick amongst the datasets used for the group assignment V if necessary, though other datasets of your choosing are also possible).

Each week, your task is twofold: 1) to complete, and keep a well-annotated record of, the code and generated results from these assignments; 2) to familiarise yourself with the code and the method enough such that if you are questioned on any part of the coding or modelling choices, you are able to defend your code. It is your responsibility to ensure you complete the exercises properly. At the end of each week, you will be exposed to a mock oral defence, where TAs / CLs will test your ability to defend coding and modelling choices informally, and give you feedback on how to better explain yourself verbally.

At the beginning of week 9 of the semester, on April 2nd, you will hand in a copy of your coding portfolio. The portfolio is evaluated according to clarity and execution criteria, which are covered below. It is necessary for your code to run smoothly without hiccups in order to pass the individual DE assignment – the TAs in charge of replicating your code must be able to run it top-to-bottom on their own machine, and will be in touch with you if they cannot do this. You will have one chance to correct the code to ensure replicability.

On April 4th and 5th, you will undergo your individual oral explainer session. You will be given the opportunity to nominate two of your submitted code notebooks to be the topic of the defence and CSSci staff will select one of them; you will **not** be notified of the final selection before the defence. If no nomination is made, CSSci staff will chose between all submitted notebooks. The defence will last around 15 minutes, and two examiners from the CSSci staff will be present. Your job will be to:

1. Answer a series of questions related to your code, and/or
2. Provide accurate justifications for the chosen methodology's compatibility with your data, and/or
3. Provide accurate justifications for the chosen methodology's compatibility with the goals of the analysis, and/or
4. Comment on the results of your DE effort.

Deadlines

Friday, February 23rd 2024 (2 to 5 p.m.)	**Mock Explainer Session 1** Optional but strongly encouraged.
Friday, March 1st 2024 (11 a.m. to 3:30 p.m.)	**Mock Explainer Session 2** Optional but strongly encouraged.
Friday, March 8th 2024 (2 to 5 p.m.)	**Mock Explainer Session 3** Optional but strongly encouraged.
Friday, March 15th 2024 (11 a.m. to 3:30 p.m.)	**Mock Explainer Session 4** Optional but strongly encouraged.
Friday, March 22nd 2024 (2 to 5 p.m.)	**Mock Explainer Session 5** Optional but strongly encouraged.
Tuesday, April 02, 2024 (5 p.m.)	**Final code submission** Final submission via CodeGrade on Canvas.
Tuesday, April 02, 2024 (5 p.m.)	**Code notebook nomination** Details provided by your Core Lecturer.
Thursday, April 04 and 05, 2024 Time by appointment.	**Code Explainer Session** Details provided by your Core Lecturer.

If you have an unavoidable conflict on April 04 and/or 05, notify your Core Lecturer as soon as possible.

Materials Provided to Students

After each DE practical from weeks 3 to 7 of the semester, students will be provided with a coding assignment. At the end of the assignment, a minimal set of solutions with a standard dataset will be provided for students to take inspiration in crafting their own solutions.

Assessment Criteria

Oral Explainer Session

- **Mastery of Code and Methods:**
  Ability to clearly explain the code, its structure, and the underlying methodology. This includes the rationale behind specific coding and modeling choices.
- **Problem-Solving Skills:**
  Ability to answer questions and defend coding choices effectively, showcasing problem-solving skills and adaptability in thinking.
- **Communication Skills:**
  Clarity, coherence, and effectiveness of verbal communication. This includes the ability to explain complex technical concepts in an understandable manner.

There are questions about which workbooks have to be handed in, and which have to be discussed for the DE individual grade. Here is a clarification:

In our DE workshops and lectures since week 3 we have been discussing a series of methods. These methods are the following: Nominal Data Models (Logistic Regression); Count-Data models (Log-Linear and Poisson Regression); Regression trees and Forests; Time Series Analysis .

Each of these methods has relevant workbooks associated with it:

- For Nominal Data Models (Logistic Regression) --> WB 1 (fitting), WB 2 (interpretation), WB 3 (evaluation);
- For Count Data Models (Log-Linear and Poisson Regression) --> WB 4 and WB 0.b (the evaluation workbook from camp week);
- For Regression Trees and Forests --> WB 5 and WB 6 (you will get these on the week of March 11th) and either WB 0.b (evaluation workbook if you choose to use trees for regression) or WB 3 (evaluation workbook if you choose to use trees for classification);
- For Time Series Analysis --> WB 7, WB 8 (you will get these on the week of March 18th) and WB 0.b (evaluation workbook for regression).

For your Code Portfolio, you have to submit a version of each workbook applied to a dataset of your choice.

For your Explainer Session oral defence, amongst these four methods, you pick two - e.g. you pick Nominal Data and Count Data. Then you have to be ready to defend the workbooks associated with these two methods.

Note that when you choose the methods you want to defend in the Explainer Session, you should ensure the relevant evaluation workbook is presented along with that method. To reduce the workload a little bit, we have decided to not require all four evaluation workbooks in your Code Portfolio; you only need to submit the two associated with the methods you will nominate for your graded Explainer Session. See the table below for the specific workbooks.

In the DE Assignment Description, what we have in lecture/workshops been calling "methods" are referred to as "notebooks." We recognize that this has caused some confusion and hope that this clarifies things for students.

METHOD **(aka Notebook)**	**CODE PORTFOLIO** **(submit ALL via Canvas)**	**EXPLAINER SESSION** **(pick two Methods for oral defence)**
Nominal Data Models (Logistic Regression)	WB 1 (fitting) WB 2 (interpretation)	WB 1 (fitting) WB 2 (interpretation) WB 3 (evaluation)
Count Data models (Log-Linear and Poisson Regression)	WB 4	WB 4 WB 0.b (the evaluation workbook for regression; from camp week)
Regression Trees and Forests (you will get these on the week of March 11th)	WB 5 WB 6	WB 5 WB 6 Select **one** of: - WB 0.b (evaluation workbook if you choose to use trees for regression) - WB 3 (evaluation workbook if you choose to use trees for classification)
Time Series Analysis (you will get these on the week of March 18th)	WB 7 WB 8	WB 7 WB 8 WB 0.b (evaluation workbook for regression; from camp week).
Evaluation workbook	The two evaluation workbooks associated with the methods you nominate for the Explainer Session. Select **two** of the following: - WB 3 for Nominal Data - WB 0.b for Count Data - WB 0.b for Regression Trees and Forests - WB 3 for Classification Trees and Forests - WB 0.b for Time Series	

Coding Portfolio

- **Completeness and Accuracy:**
  The extent to which the code fulfills the given assignments and correctly applies the demonstrated methods.
- **Documentation and Annotation:**
  Quality of comments and documentation within the code. This includes clear explanations of the logic, purpose of different code sections, and documentation of the methods and data sources used.
- **Innovation and Creativity:**
  Assessment of the uniqueness and creativity in approaching the problem and data. This could involve innovative use of algorithms, unique data manipulation, or creative problem-solving strategies.
- **Replicability and Robustness:**
  The ease with which the code can be replicated and run on a different machine. This includes the absence of hard-coded paths or system-specific dependencies, and overall robustness of the code against different datasets.

Learning Objectives

Students are able to:

- Demonstrate a thorough understanding of machine learning and other digital techniques for prediction.
- Understand how social biases manifest in digital data and predictive algorithms.
- Detect social biases in digital data and predictive algorithms.
- Implement appropriate social bias mitigation techniques.