

Inf2- Foundations of Data Science

Coursework 2 (CW2)

Critical Evaluation of a Data Science Study

Released: Monday 11 January 2021

Deadline for submission: Monday 25 January at 4pm on Learn (via the Turnitin Link).

As per the Informatics [Late Coursework & Extensions](#) policy, late coursework without an authorised extension will be recorded as late and the following penalties will apply:

- 5 percentage points will be deducted for every calendar day or part thereof it is late, up to a maximum of 7 calendar days.
- If you have not submitted coursework within these 7 days, a mark of zero will be recorded.

It is very important that you read and follow the instructions below to the letter: you will be deducted marks for not adhering to the advice below.

Good Scholarly Practice

Please remember the University requirement as regards all assessed work for credit. Details about this can be found at: <http://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct>

Specifically, this coursework must be your own work. We want you to be able to discuss the class material with each other, but the coursework you submit must be your own work. You are free to form study groups and discuss the concepts related to, and the high-level approach to the coursework. You may never share write-ups.

General Instructions

- This assignment accounts for 20% of the mark for this course and is graded based on your submission. The actual assignment is marked out of 100. The criteria on which you will be judged include accuracy, depth of understanding, critical insight. Each question is marked as either **Basic, Intermediate or Advanced, depending on their difficulty**, and is graded on a 5 point scale (inadequate, poor, fair, good, excellent). We will map the grades for each question onto an overall mark out of 100. *Roughly speaking*, you will get lower marks by doing well on the basic questions but not on the intermediate or advanced questions. To get a mark in the 60s you will need to be good or excellent on some intermediate questions, and good on at least one advanced question. To get a mark in the 70s you need to be excellent on all the basic questions, and excellent at least one intermediate question, and so on.
- Your submission should be in the form of a PDF document submitted using the Turnitin link in Learn.

Your task

In CW2, you will critically evaluate data-driven methods and claims from case studies, in order to identify and discuss how the stated conclusions are warranted given evidence provided. You will also

evaluate how well an accompanying media report covers the research. We will set out the structure for your evaluation. It's worth noting that this sort of structured evaluation of research and reporting happens in the real world, for example, see the "NHS News" summaries: <https://www.nhs.uk/news/>

We have provided you with several possible studies to choose from. Each study includes a report in a popular media outlet, as well as the scientific paper that the report was based on.

Media report	Scientific Paper
Social sickness: How Twitter can tell you (up to eight days in advance) when you are going to get ill (<i>Daily Mail</i> , 30 July 2012)	Sadilek A, Kautz H and Silenzio V (2012). Predicting Disease Transmission from Geo-Tagged Micro-Blog Data <i>Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence</i>
Coronavirus research reveals a way to predict infection without a test (<i>The Conversation</i> , 11 May 2020) Coronavirus: Five things a Covid-19 symptom-tracking app tells us (<i>BBC website</i> , 23 May 2020)	Menni C, Valdes, AM, Freidin MB. et al. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19 <i>Nat Med</i> 26 , 1037–1040. doi: 10.1038/s41591-020-0916-2
Teenagers who dislike their own body are THREE TIMES more likely to be depressed as adults, study finds (<i>Daily Mail</i> , 20 December 2020)	Bornioli A, Lewis-Smith H, Slater A, et al (2020). Body dissatisfaction predicts the onset of depression among adolescent females and males: a prospective study . <i>J Epidemiol Community Health</i> doi: 10.1136/jech-2019-213033

Choose one of the studies. Use the template provided to supply answers to the following questions.

- The following questions relate to the scientific paper of your chosen study. Note that the average length of a sentence is 15-20 words, the average length of a paragraph has between 3 to 6 sentences.
 - [Basic] What is the scientific goal of the study? Write one to three sentences
 - [Basic] What is the type of study (e.g., randomized trial, prediction model)? Include a brief explanation why it falls onto that category. Write one to three sentences
 - [Advanced] What are the hypotheses of the study? Write one paragraph (3-6 sentences).
 - [Advanced] What is the methodology of the study and what are the results (how do the researchers go about testing the hypothesis, and what is the outcome? Write two or three paragraphs.
 - [Intermediate] What are the statistical methods used in the study and how are they applied to the data? Write one or two paragraphs.
 - [Basic] What are the stated conclusions of the study? Write one paragraph.
 - [Advanced] Provide a critical discussion of the paper. How important or impactful is it? Are there any obvious errors or potential flaws? Write one or two paragraphs.
 - [Intermediate] Are there any ethical implications of the study? If none, please state so; if yes, how well do the authors relate to these implications? Write one paragraph.
- The following questions relate to the media report of your chosen study:

- a. [Basic] Provide a brief summary of the report. Write up to three sentences.
- b. [Advanced] How accurately did the report summarize the study? Write one or two paragraphs.

Advice

- When you first read the studies, you will probably find that you don't understand several concepts in the papers. Don't worry; this is normal. Part of the skill of understanding a study is to identify parts you don't understand, and see what you can understand from the rest of the paper.
- Two of the papers discuss confidence intervals, hypothesis testing and logistic regression. The lectures on these topics will be released in Weeks 1 and 2. We will try to release drafts of the lecture notes as soon as possible.
- ["How to read a paper" by Michael Mitzenmacher](#) gives very good advice on reading a paper *critically*.