

Cross-disciplinary Higher Education of Data Science Beyond the Computer Science Student

Responses to reviewers

The author would like to thank the reviewers for their comments. The comments of reviewers are addressed and below you can find our responses to each individual comment.

Reviewer: 1

“My biggest concern with this paper is: I am not sure what exactly is its contribution. I like many pieces of this paper, but I don’t know how do they come together to help the research question tackled here, or the new position advocated. Namely, is the position of this paper that there should be (more) interdisciplinary courses that rely on data science (I am not sure how novel would this be)? Or is the contribution in some of the design choices presented e.g. in section 4? Or is the contribution in understanding the relation of some decisions to the theories presented in section 2? I think this is not clear in the paper currently. Some of my points 1.1-1.5 might be resolved once this point is made clearer.”

The author clarified the contributions of the paper at the introduction as follows: “The contributions of this paper are (i) the analysis of effects and implications by design choices made to address a cross-discipline approach to data science as well as (ii) lessons learnt after teaching a cross-disciplinary data science course for 3 years at a top-class university. ”

“Related to this, I find the lessons learned in section 7 unclear. Namely, the lessons ”software tools” and ”data requirements” do not seem to be really learned from the course implementation, but are rather design decisions of the course. Some of the other lessons follow from the course, but then it is important to keep focus and mainly discuss them if they help the position presented in this paper. Also, the author might want to link these lessons to the results in section 6. I am not sure what is the role of the theories in section 2. I was expecting that these are used in section 7 to explain some of the outcomes from the course.”

The author improved the presentation of the lessons to bring more in the foreground take away messages for the reader. Links to results in Section 6 as well as to the learning theories are added.

“I am missing related work that is on the same level as this paper: have other people created and discussed similar cross-disciplinary courses? How does this course differ?”

A new section “Comparison of Related Work on Data Science Education” has been added outlining research on data science education.

“The tables presented in this paper are clear and informative. But I am unsure what is their relation to the position in this paper (e.g. why does it matter that there are PhDs following the course)?”

The tables contribute to the completeness of information provided about the course. They also provide evidence about the diversity of the students’ background.

“I am wondering whether this course is truly cross-disciplinary. Namely, it seems to me that the

course teaches people (from more backgrounds than just CS) to deal with big data. So this is great, and through the applications there is a frame for the course work to be applied to improve existing technology. Still, this link seems one-directional to me, from the data science domain towards application areas where the data science seems logical to apply. But for "true" cross-disciplinarity, I would expect more discussion on the application domain side, namely, looking at real, important problems in various domains that need to be solved. If the application domains are instead "toy" cases, then this course does not differ much from any data science course, because I can imagine one always works on a dataset that represents a certain domain. I believe that the course designers/students have looked into finding meaningful application problems, but I think this could use some explanation in the paper."

The actual cross-disciplinary nature of the course is better outlined in the last paragraph of Section 4 by stretching the development of domain knowledge.

"Are the answers given by the interviewees 1 to 5 explainable/expected based on their background? I would be interested if this could teach us a lesson and suggest whether some backgrounds fit better the current form of the course than others."

A few remarks on this aspect are added in the lessons learnt.

"I find the paragraph on page 4 "ETH Zurich offers.." too abstract and not too informative. Could it be made more specific (this might be just a matter of re-wording some of the sentences)?"

Sentences of this paragraph are rephrased.

"Section 4 needs a bit more focus and consistency - there is a lot of explanation of the first aspect of the course, while less/different detail is given for instance on the Machine learning part. Additionally, is "social media" not falling better under the "applications" aspect?"

Section 4 is improved to reflect reviewer's remarks.

"Typos and grammar issues: "discipline"->"disciplines" (page 1), "learnign" (page 2), "exlusively" (page 3), "should encounter for this diversity and has the capacity" -> "should encounter for this diversity because it has the capacity"? (page 3), "students attended" -> "students that attended" (page 4), "Table 2 illustrates..." -> "Table 2 illustrates the semester status of the students who have participated in the course" (page 5), "Step 6" -> "Step 7", "Step 7" -> "Step 8" (page 8), "with 5 students attended" -> "with 5 students that attended" (page 10), "what data mean" (page 13)."

All of these issues are fixed.

Reviewer: 2

"In the introduction, i would expect more background, for instance a comparison with other post-grade training courses for data science or disciplines of similar cross-disciplinary characteristics such as Bioinformatics."

The author added some references and a new section reviewing related courses and programs on

data science education.

“In page 2, the following claim the authors backed with references: “Given the evident lack of plurality and interest for data scientists in the job market, ...” is very surprising to me, above all with an increasing number of news like [1] and [2] that state the contrary. It would be beneficial for the argumentation to discuss it in the context of these opposite analyses.”

The author rewrote this remark to make the statement clearer.

“In section 3, they state the lecturers of the course are not cross-disciplinary, although they all have experience in multi-disciplinary research. May this fact, the lack of presence of cross-disciplinary lecturers, have effect on the learning results of the course? Do they have plans to evaluate that?. I am also wondering if there is a reason why the course schedules during spring semester.”

The lecturers’ background is evaluated in Question 4 of the interviews. It is further added that to improve the cross-disciplinarity outcome of the course, lecturers of different discipline can teach part of applications in the future, which can allow an evaluation of the learning outcome. Evaluation is indicated as part of future work. There is no special reason for running the course during spring semester.

“In table 1, figures suggest an increasing lose of interest by life sciences students, could you comment on it? Is the presented course following the Bologna process [3]?”

With the small number of total students during the first year, it is not possible to draw conclusions at this level of detail and that is why the author does not comment on isolated cases. The Bologna process at ETH is usually introduced at a higher level, i.e. study programs (<http://e-collection.library.ethz.ch/eserv/eth:27010/eth-27010-01.pdf>).

“Regarding some observations made by students about course evaluation in section 6, do the authors think that setting up IT technologies such as online forums, could improve and foster communication among students and tutors, and transmission of background knowledge and skills?”

The author has added some discussion and references about the role that gamification could play to foster communication, participation and richer interactions.

“Regarding lessons learnt exposed in section 7, in line to cultivate critical thinking and constructive doubt, what the authors think about to include an evaluation made by counterpart students on each research project as a part of the final course evaluation? and finally, do they plans to evaluate the real impact of the course in the job market?”

Students participate in the final project presentations and can challenge their classmates with questions. The evaluation of the course in the job market is part of future work.

“On the other hand, there are some other presentation issues that could be poulish: the paper is too long, repetitive sometimes (e.g. page 2, introduction explains twice what this paper is about). Some parts seem to be written in a hurry with some typos, and lists of references do not have a consistent logical order.”

- “ - p2: typo 'learnign'
- p2: typo 'costructivism'
- p4: 'they concern state-of-the-art' -> 'they concern on state-of-the-art'
- Tables 1 and 2: what does 'n' denotes?
- p5: redundant sentence: 'Sections 4-6 illustrate the content and research projects of the course as well as the course evaluation and students' feedback.'
- p5: 'combining the behavioral and design science research strategies.' - i do not understant the meaning of 'behavioral' in this sentence.
- p7: in 'adjusted in an education context' -> replace by 'educational'
- p7: 'results written and orally' -> 'results in writing and orally'
- p8: 'stands the selection' -> 'stands for the selection'
- p8: 'in their illustrations' -> 'in their presentations'
- p8: 'high-quality illustration' -> 'high-quality presentation'
- p8: guide -> guideline
- p8: in the text: step 5 -> step 6
- p8: in the text: step 6 -> step 7
- p8: in the text: step 7 -> step 8
- p9: revision of verb tenses: 'this project conducted' -> 'this project was conducted'; involves -> involved; the project is -> the project was:
- p9: typo 'ran be' -> 'ran being'
- p10: 'it proved not straightforward' -> ..not to be..
- p10: 'two official evaluation' -> evaluations
- p10: 'the general satisfactions' -> satisfaction
- p10: agenda question 3 'what was ... factors' -> factor
- p11: Table 4 caption: Background -> Educational background
- section 6: could be reduced placing interview answers as Supplementary material.
- p12: i think this sentence should be vice-versa like: 'The feedback suggests that working/lab sessions during the class may motivate further the non-computer scientists to improve their knowledge as well as the computer scientists to practice their skills during the course.'
- p14: 'participating citizen' -> participatory citizen
- The access to the interview survey described in section 6 is not FAIR. ”

The above concerns are addressed.

“Indeed this is a relevant topic for the data science community, how to train data science in regard to the job market needs. I think a paper like this is appropriate as a position paper for a data science journal, and i would like that this boosts and active and necessary debate in the community. However, i would expect a bolder claim on the societal impact of training data science for a broader scope of students. This is already stated in the paper but maybe the authors could highlight their vision on the important position of data science in a digital society.

To sum up, the paper raises a novel topic of important societal impact: education of data science for future professionals, and opens an interesting discussion on training data science for students from any educational background. But the papers suffers from verbosity and some presentation issues. It needs

careful reading and polishing.”

The author made several and very careful revisions of the paper that capture these comments and improve the overall quality.

Reviewer: 3

“first of all, the position of the author in this position paper is, in my opinion, obfuscated by their experience. While the latter is extremely valuable to support the authors opinions and findings, I would highlight these more;”

The author made several changes throughout the paper to bring more in the foreground the position of the paper.

“the paper is centered around the course that the author ran together with his colleagues. This direct experience is important, but I miss a bit of comparison with other similar courses. For example minor courses and other courses dedicated to non-computer scientists from other institutes;”

A new section is added with comparison to related work on data science education.

“regarding the multidisciplinary aspect of data science and the challenge of balancing the computer science point of view with those from other disciplines, I would like to see a comparison with the Web Science curricula. Also in the Web Science community, a comparable discussion has been promoted, given that also Web Science is a multidisciplinary and emerging discipline. I wonder how similar and how diverse the two resulting disciplines (and curricula) are;”

This is definitely an interesting point which is now mentioned and referenced in the paper as a direction of future work for the course evaluation.

“in the introduction, the author states Given the evident lack of plurality and interest for data scientists in the job market. It seems to me that while there may be lack of plurality, the job market is increasingly interested in data scientists (but I may have a wrong perception about this). Nevertheless, I wonder if the course described contributes to tackling this problem. Given that the course has been running for three years, would it be possible to evaluate if teaching data science in this manner to non-computer scientists actually contributed to their career development?”

This statement is rewritten to clearer convey the point. Moreover, the evaluation of the career development is an interesting aspect and deserves its own research study. Therefore it is mentioned as part of future work.

“Sections 2 and 3 would probably benefit from a classification of the course with respect to Bloom’s and Dee Fink’s taxonomies;”

The learning objectives of the course are illustrated in a new figure according to the Bloom’s taxonomy.

“lastly, this paper focuses on teaching data science in a multidisciplinary context, where the multidisciplinary aspect comes from the students’ background. I wonder if (and how) the lessons learnt can be used also to bring multidisciplinary to the courses taught in the context of core data science, like the big data, data mining and machine learning courses mentioned: would it be possible/advisable to bring more multidisciplinary aspects also in these courses (or in the curricula they belong to)?”

This very interesting aspect is mentioned as part of evaluation in future work.

“Note that several groundbreaking discoveries in the areas of complex networks and biology [...]”: please, cite (at least some of) them. [last paragraph of Section 2]: For instance [...] for instance. Please, revise.”

These issues are now addressed.