

Response to reviewers: “Kaggle-in-class Data Challenges Can Boost Student Learning”

We would like to thank the reviewers and the Associate Editor for their thoughtful second review of the manuscript. The reviews have further improved the manuscript.

Please find below the *referees’ comments (italics)* and our response (blue).

Reviewer: 1

Comments to the Author

I think the previous concerns were addressed and the paper would be useful for teachers.

1. *Section 2: Having more information on the data used in the competitions should be provided, how many variables? How many observations in the training and test data?*

Complementary document was added with more info about the datasets.

2. *The author might want to call the undergraduate students used a comparison group not a control group. Control to me implies random assignment.*

Changed.

3. *I would suggest adding a section in the discussion for study limitations. Some are addressed in other sections but a specific paragraph or two on limitations would be helpful.*

The summary of study limitations has been added to the discussion section.

4. *It would be useful to have summary statistics for the groups for an exam they took early in the course before the competition, and summary statistics on the percentage correct overall on the final exam for the different groups, in addition to your calculated performance scores.*

Additional summary statistics have been added to section 2.3 Participants.

5. *Finally, are the data you used in these competitions available to teachers? If yes where can they be found or how can teachers get them from you? I know you put in some randomness but having the raw data available by request for instructors would be helpful.*

We are happy to share the data by request. This is now reflected in a comment in the “Teachers’ corner” section.

6. *minor comment: page 7 line 29, missing a space in the effects*

Fixed.

Reviewer: 2

The authors have successfully addressed many of the observations raised by this referee during the review of the original submission. Below are minor comments that the authors should consider for the further improvement of their manuscript.

Minor comments

1. Page 2: From this sentence it might seem that the mentioned algorithms were created as a result of Kaggle competitions, which obviously is not the case. Reword appropriately
"Occasionally new algorithms have emerged, for example, deep neural networks (Hinton & Dahl 2012) and XGBoost (Chen & Guestrin 2016)."

The sentence was reworded.

2. Page 3: competition is not an assessment method but an active learning method, when it comes to sentence "A competition, like any other **assessment method** has its advantages and disadvantages."

"assessment method" was changed into "active learning method, that used for assessment".

3. Page 3: Space is missing between "the" and "effects" in "This paper contributes to this call by offering statistical analysis of **theeffects** on learning of classroom data competitions."

Fixed.

4. Page 4: "'gold standard" is not quoted appropriately

Fixed.

5. Page 4: Change the word "mixing" to "ensemble" or "averaging" in "advantages of group work and **mixing** models"

Changed to "ensemble models".

6. Page 5: From the argument that "In the years prior to this experiment, the undergraduate scores on the final exam are indistinguishable from those of the graduate students", it sounds like the finals of undergraduates and graduates are of the same difficulty, which does not sound correct. In addition, I am still unconvinced that the consideration of undergraduate cohort as the control group was well-thought.
Authors need to elaborate more carefully on this point.

To justify our choice of undergraduate students as a control group for graduate students we added some further discussion and summary statistics to section 2.3 Participants.

7. Page 14: replace "avoid **cheating**" with "avoid plagiarism and use of unauthorized assistance"

Done.

8. *Page 14: Mention that "Categorization Accuracy" means percent of correct classifications.*

Changed to: "Categorization Accuracy", the percent of correct classifications, is reasonable."

9. *Page 14: Reword the sentence to read better: "If you are running a regression challenge, then "Root Mean Squared Error (RMSE)"."*

Reworded to: "If you are running a regression challenge, then the "Root Mean Squared Error (RMSE)" is a good choice."

10. *Sentence "Students are often motivated to consult with the instructor about why their model is underperforming, or what other approaches might produce better results." on page 16 contradicts the sentence "The competition needs to run without any intervention from the instructor." on page 15*

We didn't consider academic discussion with students about covered learning material as an intervention. But we can see the point and therefore add the following two sentences on page 15 immediately after the sentence "The competition needs to run without any intervention from the instructor.":

"Exception is of course an academic discussion motivated by the competition between the teaching team and the students. For example, a discussion about different models, their advantages and limitations.

Associate Editor

Comments to the Author:

See minor suggestions/corrections below:

1. *Page 3 Line 12: Maybe use "However" instead of "But"*

The suggested change done.

2. *Page 3 Line 29: Separate "the effects"*

Fixed.

3. *Page 3 Line 54: Remove comma*

Done.

4. *Page 4 Line 49: "Another motivation" sentence is bit awkward/unclear*

Changed to: "Another reason for this stratagem was the university policy, requiring a strategy to assign students individually in group assignments. When the team members develop the model together, it is quite difficult to accurately assess the individual contribution of each student. The individual submissions helped to encourage each student to engage in the modelling process. In addition, it helped to assess the individual component of the final score for the competition."

5. *The use of ETC 2420/5242 and MAST90083 as codes for the two classes was not completely intuitive. I did need to flip back to the original explanation of what those class codes referred to several times. I realize those may be the college course numbers, but it might be clearer to use a different way to identify the two courses (especially the ETC course) throughout the paper.*

Notation changed:

MAST90083: **CSDM** (stands for Computational statistics and data mining)

ETC 2420/5242: **ST** (stands for Statistical thinking)

ETC 2420: **ST-UG** (stands for Statistical thinking-undergraduate students)

ETC 5242: **ST-PG** (stands for Statistical thinking-postgraduate students)

We hope that these notations will be more intuitive for the readers.

6. *This was most noticeable on the scatterplots on page 12 where you added R and C as part of the code. Maybe a legend would help here? I realize you may be limited by JSE formatting guidelines, but it might make the scatterplots a bit clearer.*

We added the subtitle in plots "Classification" instead of "C" and "Regression" instead of "R"

7. *Page 16 line 42: I am not sure you need to use the term "anecdotally". To me that implies that you are just sharing your informal observations of your classes. There was a survey conducted and analyzed, even if the survey design had issues (which you addressed as part of the survey section).*

The word "anecdotally" was removed.