Harvard Data Science Review • Issue 3.1, Winter 2021

Project-Based Learning via Competition for Data Science Students

Philip L. H. Yu¹ Wai Keung Li¹

¹Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong

Published on: Feb 25, 2021

DOI: <https://doi.org/10.1162/99608f92.44f54f00>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

We would like to congratulate Professors Eric Kolaczyk, Haviland Wright, and Masanao Yajima for their very interesting and impressive work (“Statistics Practicum: Placing ‘Practice’ at the Center of Data Science Education,” this issue). They have addressed the important issue of theory versus practice in the training of a statistician/data scientist by placing a practicum course at the center of a data science program, and shared their valuable experience in their M.S. in Statistical Practice (MSSP) program. Below we share some of our own experiences that resonate with the theme of Kolaczyk et al. (2021).

Data Science Projects

As Kolaczyk et al. (2021) detail, one of the key components of their MSSP program is a two-semester statistics practicum course in which students are involved in a number of external-partner projects and consulting projects for the university community. We concur with the authors on the importance of embedding a substantial practicum in the curriculum. However, in our situation, industrial consultation is less developed compared with that at Boston University. Nevertheless, it is relatively easy to collect research projects within the university. Many of these are, from the point of view of a data scientist, applications oriented and interdisciplinary in nature. When we delivered a talk titled “Statistical Curricula Development at the University of Hong Kong” in an invited paper session on “Developing Undergraduate Curricula for Statistical Workplaces Now and Future” at the World Statistics Congress in 2013, we mentioned several distinctive aspects of the development, including the need to integrate statistical thinking and reasoning, massaging data, interdisciplinary inquiry, and research-based teaching and learning (Yu & Li, 2013). Statistics in action has been one of the major core values of reforming the undergraduate curriculum.

In 2017, when we designed the curriculum of the Master of Data Science (MDASC) program, an interdisciplinary program jointly offered by Departments of Statistics and Actuarial Science, and Computer Science of the University of Hong Kong, we developed a two-semester data science project as a core course. The projects are mainly research-based projects given by colleagues from the two departments, and a number of them may involve external partners as supervisors. For example, in a series of projects related to the detection of suicide-at-risk messages from social media, researchers from the Hong Kong Jockey Club Centre for Suicide Research and Prevention provided expert advice for the project.

Project students not only have opportunities to apply theory into practice, they could also make a meaningful contribution to our society with the knowledge and skills gained throughout their projects. More importantly, students may deepen their understanding about the entire process of data analysis and appreciate the importance of data collection and preparation to the success of the whole project.

Innovative Data Mining Application Awards

As an important and integral component of data science, a data mining course has been offered since 2003 in the Master of Statistics program at the University of Hong Kong and recently in the MDASC program. To let

students experience the excitement and taste of success from learning, a data mining competition has been organized and become a flagship event every year. In the project, students form groups to apply the knowledge learned in the course in mining a real data set from their workplace, international data mining competitions, or elsewhere. Each group should write a project proposal, find the necessary data, analyze the data, and present their findings. The top three teams are awarded the Innovative Data Mining Application Award sponsored by an industry partner. We organized a prize presentation ceremony jointly with the sponsor. Some clients of the sponsor were also invited to the presentations. This really helps boost the job-seeking success rate of the students. We concur with the authors that assessment of practicum success “cannot be achieved only by the judgment of academia faculty.” In project presentations, students tended to present in a rather technical way. To help students improve their communication and presentation skills in their future workplace, the top-scoring project teams were arranged to present the projects to high school students who can act as their ‘boss,’ but who do not have much knowledge in data mining. Feedback from high school teachers involved indicates that their students were amazed with the power of data analytics in solving complex problems in our daily life. Data analytics are more than just plotting charts and graphs in spreadsheets. Every year, around a hundred high school students were engaged through the presentation period and actively involved in the question-and-answer session. The presentations serve the dual purpose of helping the master’s students in sharpening their presentation skills and promoting data science education to the high school students. Finally, we published a booklet that summarizes the awarded projects to recognize the enthusiasm of the students in applying the knowledge and experience gained from the course to explore new frontiers of innovation. Over the years, we collaborated with students who worked in an applied project to publish their works in journals, newsletters, magazines, and newspapers (see Li & Yu, 2015; Qian & Yu, 2019; Zhang & Li, 2019).

Challenges Ahead

Due to the COVID-19 pandemic, formal classroom teaching and learning in higher education is conducted online in Hong Kong and worldwide. Online learning usually includes both synchronous and asynchronous learning environments with internet access to increase flexibility and foster lifelong learning (Dhawan, 2020). Very often during live online classes, students tend to keep their cameras off due to various reasons such as increased anxiety and stress, privacy, internet access problems, and so on. Both supervisors and students of the proposed practicum course might have been experiencing unprecedented challenges like scheduling online meetings with different parties, including supervisors, clients, and external partners, tackling problems raised by students in hands-on practices of computer programming, and engaging with students through distance learning. Recently, gamification has shown to be an effective learning strategy to create engaging experiences in online learning (Zainuddin et al., 2020). The pandemic is creating a ‘new normal’ in education and accelerating the development of innovative teaching and learning technologies so as to achieve a responsive, ethical, and humane approach to data science education.

Disclosure Statement

Philip L. H. Yu and Wai Keung Li have no financial or non-financial disclosures to share for this article.

References

- Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>
- Kolaczyk, E., Wright, H., & Yajima, M. (2021). Statistics practicum: Placing ‘practice’ at the center of data science. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.2d65fc70>
- Li, Y., & Yu, P. L. H. (2015). Visualizing big ranking data. *Bulletin of Hong Kong Statistical Society*, 37(1), 9–13. https://www.hkss.org.hk/images/book/HKSSBulletin_V37_1_Mar2015.pdf
- Qian, Z., & Yu, P. L. H. (2019). Weighted distance-based models for ranking data using the R package. *Journal of Statistical Software*, 90(5), 1–31. <https://doi.org/10.18637/jss.v090.i05>
- Yu, P. L. H., & Li, W. K. (2013). *Statistical curricula development at the University of Hong Kong* [Conference session]. World Statistics Congress, Hong Kong, China, August 26, 2013. <http://2013.isiproceedings.org/Files/IPS060-P3-A.pdf>
- Zainuddin, Z., Shujahat, M., Haruna, H., & Chu, S. K. W. (2020). The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system. *Computers & Education*, 145, Article 103729. <https://doi.org/10.1016/j.compedu.2019.103729>
- Zhang, Z., & Li, W. K. (2019). An experiment on autoregressive and threshold autoregressive models with non-Gaussian error with application to realized volatility. *Economies*, 7(2), Article 58. <https://doi.org/10.3390/economies7020058>
-

©2021 Philip L. H. Yu and Wai Keung Li. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.