Teaching Basics of Data Science to non-STEM undergraduates. A Hybrid Learning Approach

Ilya Musabirov[1], Alina Bakhitova[1] Paul Okopny[2], Stanislav Pozdniakov[1], and Alena Suvorova[1]

[1] National Research University Higher School of Economics, Russia
imusabirov@hse.ru
[2] Uppsala University, Sweden
paul.okopny@gmail.com

Abstract. In this report, we describe the first experience of applying a blended learning approach to a Data Science Minor curriculum at St.Petersburg campus of Higher School of Economics. We employed a non-obligatory MOOC-based course in the educational process, which allowed us to enhance the learning experience of non-STEM students.

Keywords: blended learning, data science, non-STEM students

1 Introduction

In this report, we describe the experience of applying a blended learning approach to a Data Science Minor curriculum at St.Petersburg campus of Higher School of Economics.

The demand for data analysis skills is rapidly growing, and more non-STEM disciplines are moving into the interconnection with Data Science, finding it a valuable and useful experience [1]. The nature of the subject suggests a blended learning approach. Studies have shown that the inclusion of blended learning activities in the student experience can help to reduce student attrition, and is positively related to examination performance [2].

Data Science is a minor specialization, which undergraduate students can choose to study for two years (2nd and 3rd year of a bachelor program). The specialization unites students from different non-STEM programs and departments, ranging from Economics to Oriental Studies. Students study different methods and techniques related to Data Science, Text Mining and Social Network Analysis. The first cohort of students started their studies in September 2015. The second cohort started one year later in September 2016.

In addition to traditional face-to-face classroom settings with teachers and teaching assistants explaining materials, the educational process on the Data Science Minor is supported by the virtual learning environment (VLE). The VLE is a web-based software system which consists of RStudio Server IDE and a Q&A forum. The web-based nature of the VLE allows students to access the same working environment in and outside the class.

2 First Year Experience

Data Science Minor students at St. Petersburg have substantially different backgrounds. Overall, students from 10 educational programs are enrolled in the Minor: Oriental Studies (14 students from the first and 5 from the second cohorts), Public Administration (3, 2), History (6, 5), Logistics (24, 38), Management (16, 40), Political Science (7, 9), Sociology (35, 23), Philology (0, 8), Economics (46, 57), and Law (5,6). In total, 349 students were enrolled in 2015-2016.

During the first year after the launch of the Data Science Minor, we experienced several major issues:

- Skill disparity among students, as some students, especially who have previous practice in programming and IT (e.g. have more Statistics and IT-related courses), tend to grasp new material at an instant, while others may struggle through unfamiliar topics. This disparity leads to two consequences. First, students who grasp new material faster than others become bored with classes which affect their motivation to learn. Second, students with less experience often fail to communicate their problems via the Q&A forum, thus receiving less help from others.
- The number of students attending Data Science Minor makes it impossible for a teacher to reach every student in person in an attempt to re-explain unclear topic.
- Manual grading in most cases becomes difficult if not impossible.

Based on our experience we decided to employ online courses as an additional way of learning, which would complement regular class-based lessons. This strategy could support both students who struggle to learn programming and advanced students as it motivates them to do more and get higher grades, as it was suggested in [3]. The online nature of that additional component could also solve issues with scaling lessons to a bigger number of students.

3 Blended learning in Data Science Minor

In September 2016, we employed an online educational platform Stepik.org as a part of the Data Science Minor curriculum. While Stepik.org is oriented mostly to open online courses, we developed and delivered a private course for our students. These changes provided students with more flexible and convenient access to course materials and the possibility to assess and track students' progress automatically.

We transferred some course materials and assignments to the platform. Additional online course materials were structured in 7 modules, each addressing different aspects of Data Analysis: R and R Markdown basics, concepts of exploratory data analysis, text mining, visualisation, data filtering and aggregation, managing dates and time. The typical module contains lessons with 7–14 exercises (multiple choice and free-response questions, coding, matching and sorting exercises), 247 exercises in total. Each module became available for students

after the corresponding topic has been covered in class by the instructor and were open until the end of the first semester and had un unlimited number of attempts. In addition, we designed special obligatory online modules that included two midterms and one final exam. Unlike the "regular" modules the exams had limited time for completion and limited attempts.

4 Lessons Learned and Current Work

Heterogeneity of the students poses significant challenges for the developers of STEM education specializations for non-STEM students.

Part of a student's success is connected with general student's educational achievements, their performance in other courses. Fig. 1 shows that students' rankings in their major (humanities and social sciences) are significantly connected with their grades on the Data Science Minor midterms, and this connection is even tighter in the case of educational programs with less Math-related or IT-related courses.

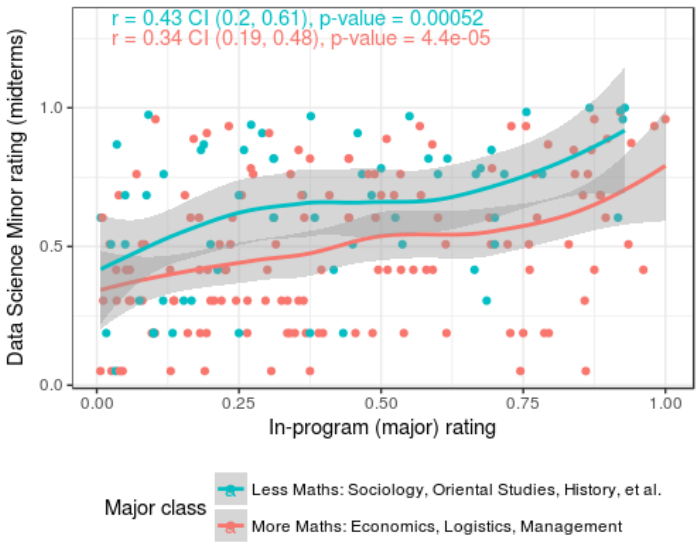

Fig. 1. Minor rating and in-program rating (1 is the top rank). The shaded areas correspond to 95% CI for LOESS regression curve. To address the differences in educational programs (majors) we separated the ten majors into two classes. The red class included majors with more Math-related and IT-related courses: Economics, Management and Logistics; the blue class included Oriental Studies, Public Administration, History, Political Science, Sociology, Philology, and Law.

The introduction of additional opportunities for online learning allows students to improve their skills or try to compensate the starting skill disparity:

the percentage of completed exercises at the online course is positively correlated with the results of in-class tests ($\rho = 0.504$, $p \leq 0.01$).

However, more detailed look shows that the difference in strategies use for the online component is significant. Fig. 2 shows the case of two different educational programs that there is a mix of students with high grades and low usage of online component and vice versa.

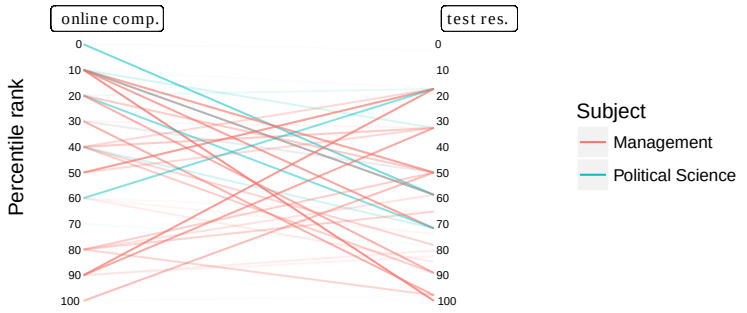

Fig. 2. Online component usage (in %) and test results (in %, 100 is successfully completed test). Each line represents student's strategy (e.g. completion of 20% assignments and 90% success on test or vise versa).

These findings highlight the importance of taking into account the skill disparity and diversity of styles by introducing adaptive exercises [4] and multidimensional models of students' skills. To explore what parts of the online experience worked and didn't work we plan to analyze students behavior in details including the data about online component usage, Q&A activity and VLE programming activity.

References

1. Van Wart, S.J.: Computer science meets social studies: Embedding cs in the study of locally grounded civic issues. In: Proceedings of the eleventh annual International Conference on International Computing Education Research, ACM (2015) 281–282
2. López-Pérez, M.V., Pérez-López, M.C., Rodríguez-Ariza, L.: Blended learning in higher education: Students perceptions and their relation to outcomes. Computers & Education **56**(3) (2011) 818–826
3. Carter, J., White, S., Fraser, K., Kurkovsky, S., McCreesh, C., Wieck, M.: Iticse 2010 working group report motivating our top students. In: Proceedings of the 2010 ITiCSE working group reports, ACM (2010) 29–47
4. Paramythis, A., Loidl-Reisinger, S.: Adaptive learning environments and e-learning standards. In: Second european conference on e-learning. Volume 1. (2003) 369–379