Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #47599

BOARD # 259: IUSE: Engaging Non-Computing Majors in Hands-on Data ScienceLearning through a Web-based Learning Platform

Dr. Xumin Liu, Rochester Institute of Technology

Xumin Liu received the PhD degree in computer science from Virginia Tech. She is currently a Professor in the Department of Computer Science at Rochester Institute of Technology. Her research interests include service computing, data science, and machine learning.

IUSE: Engaging Non-Computing Majors in Hands-on Data Science Learning through a Web-based Learning Platform

Abstract

This paper describes the design and development of a web-based Data Science Learning Platform (DSLP) aimed at making hands-on data science learning accessible to non-computing majors with little or no programming background. The platform works as middleware between users such as students or instructors, and data science libraries (in Python or R), creating an accessible lab environment. It allows students to focus on the high-level workflow of processing and analyzing data, offering varying levels of coding support to accommodate diverse programming skills. Additionally, this paper briefly presents some sample hands-on exercises of using the DSLP to analyze data and interpret the analysis results.

1 Introduction

Data science has become a crucial skill across disciplines, calling for curriculum and software tools to teach students how to derive meaningful insights from data and make domain-specific decisions[1,2,3]. However, for students from non-computing backgrounds, the steep learning curve of data science poses challenges, as they often lack the programming skills required to effectively use existing tools and libraries. Meanwhile, non-computing students are typically more focused on applying data science principles to solve real-world problems within their domains than learning how to program. Therefore, programming-intensive data science curricula are not well-suited for these students.

To address this challenge, we present an innovative educational tool, a web-based Data Science Learning Platform (DSLP), which bridges the gap between non-computing students and the technical demands of data science. Acting as middleware, the DSLP integrates with popular Python and R libraries, providing a user-friendly web-based interface for specifying data analysis tasks. It then translates these tasks into executable Python/R code and returns the results. This allows students to focus on high level data analysis workflows without becoming overwhelmed by programming complexities. Moreover, to cater for the diverse backgrounds and learning expectation of stuents, the DSLP features multiple levels of coding support, enabling students with varying levels of programming experience to engage meaningfully in the learning process, ranging from a fully guided experience, to experiment with coding independently.

In this paper, we describe the design principles and technical architecture of the DSLP, emphasizing its ability to provide hands-on data science education. We also present sample

exercises that demonstrate how users can leverage the platform to understand data science concepts, practice using Python and R libraries, and interpret the results of their analyses. The remainder of this paper is organized as follows: we discuss the design principles and technical architecture of the DSLP in Section 2, illustrate its application through sample exercises in Section 3, and conclude in Section 4.

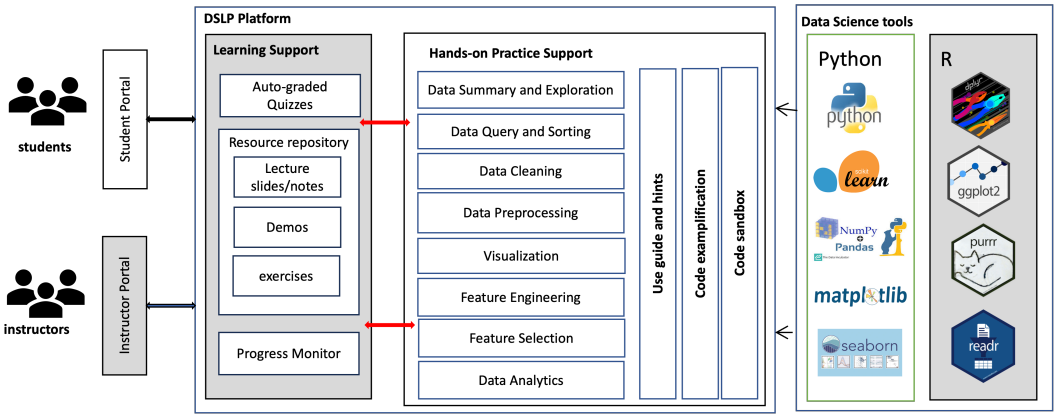

Figure 1: DSLP, a supporting platform for teaching data sciences

2 Data Science Learning Platform

The overview of the platform is shown in Figure 1. DSLP works as a middleware between users (i.e., students or instructors) and data science libraries (Python or R) to create an accessible lab environment for non-computing majors. The platform will support both Python and R data science modules due to their great popularity and comprehensive support for data science tasks. Developing a learning platform on top of these libraries will not only allow students to leverage the power of those software libraries but also help them get familiar with the libraries, teach them computational thinking, and prepare them to take further programming or data science courses. Meanwhile, it features a comprehensive learning support, including a self-assessment components and learning material repository that support both instructors students for teaching and learning. The **design principles** of the DSLP are described as follows.

D1: Provide comprehensive coverage on data science topics: To ensure that students can be exposed to various data science concepts and techniques even in an entry-level course, DSLP covers the complete list of tasks that are commensurate with the standard data-to-knowledge pipeline found in [4], including data summary, query and sorting, cleaning and preprocessing, exploration, visualization, feature engineering and selection, and data analytics. DSLP provides modules for each task where students can choose different methods and specify different options.

D2: Cater for diverse backgrounds and learning expectations: Even sitting in the same class, students could have very different programming backgrounds and learning demands. To address this, the DSLP provides (1) a user-friendly interface to allow students to operate on a dataset with no coding involved if they want to focus more on data science topics, and (2) coding support to help students learn how to write code at different levels, including a code exemplification that

generates real-time Python code for user operations and a Python sandbox that allows students to write and test their own code within the platform.

D3: Provide input for student engagement: It is well-acknowledged that student success is positively correlated to their engagement in the learning process. Monitoring student participation in class activities is important for effective teaching. Along this line, the DSLP allows to trace student interaction logs as an indicator of their engagement. performance.provided.

3 Sample Lab Assignments

We have designed lab assignments for students to use the DSLP to strengthen their understanding of the important data science concepts, including data types, data cleaning, data exploration and visualization, date preprocessing, feature engineering, feature selection, and data analytics (i.e., supervised and unsupervised machine learning models). In the Section, we briefly describe two sample assignments when students will use the DSLP to perform data visualization and feature selection, respectively.

In the first assignment, students are asked to generate a boxplot to examine if septal depth of a flower has any bearing on the type of the flower. As shown in Figure 2, students will use the visualization component to specify the type of plot and the features for the visualization. The platform will then generate both the Python code and the boxplot. Students can view the boxplot to answers questions such as: *Is there a discrepancy in the septal depths among different types of iris flowers*? Students will also be asked to revised the given Python or R code to generate a new boxplot where septal width is examined and the boxplot color is changed to red, like shown in Figure 2.

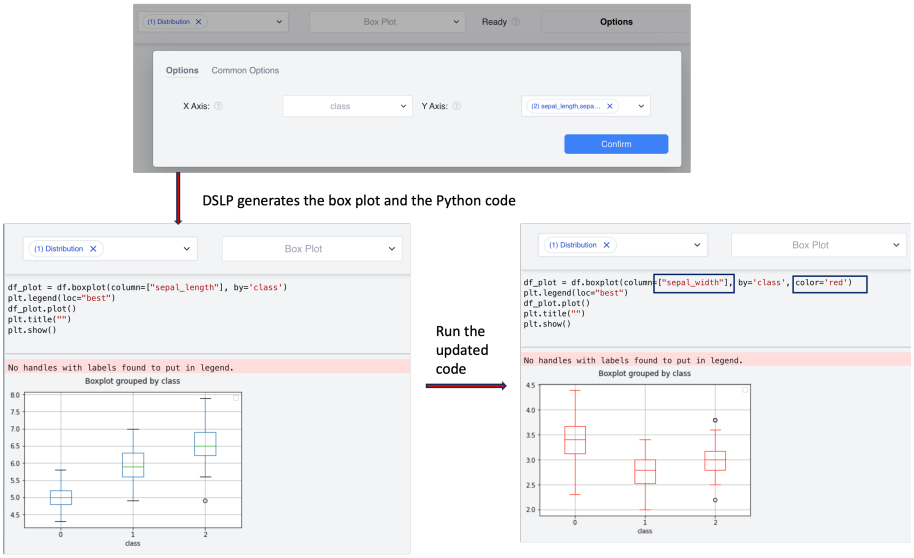

Figure 2: Data Visualization Assignment

For the second assignment, students are asked to perform feature selection on the Iris dataset in order to potentially improve the performance of supervised learning algorithms. Students will use the Feature Selection module to specify the list of features, the targeted variable, the feature

selection technique (e.g., correlation matrix), and the plot type (e.g., heatmap). As shown in Figure 3, the platform will generate the corresponding heatmap and the correlation scores to help students identify relevant or redundant features. Students will answer questions to demonstrate their understanding of the concepts and the output, such as: *which pair of features are most correlated*, *which feature of the Iris dataset is irrelevant for determining the class of a given iris flower*, and *which are the two most important features for determining the class of a given iris flower*.

Figure 3: Feature Selection Assignment

4 Conclusion

We present a web-based platform designed to provide accessible, hands-on experiences tailored for non-computing majors. The platform supports both Python and R, two of the most widely used programming languages in data science, ensuring its broader applicability. To cater to students with varying levels of programming expertise, it offers different levels of coding support, accommodating beginners while also engaging more advanced learners.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Awards #2021287 and #2336929. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Austin Cory Bart, Dennis G. Kafura, Clifford A. Shaffer, and Eli Tilevich. Reconciling the promise and pragmatics of enhancing computing pedagogy with data science. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education, SIGCSE 2018, Baltimore, MD, USA, February 21-24, 2018*, pages 1029–1034, 2018.

[2] Jeffrey S. Saltz, Neil I. Dewar, and Robert Heckman. Key concepts for a data science ethics curriculum. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education, SIGCSE 2018, Baltimore, MD, USA, February 21-24, 2018*, pages 952–957, 2018.

[3] Lillian N. Cassel, Michael Posner, Darina Dicheva, Don Goelman, Heikki Topi, and Christo Dichev. Advancing data science for students of all majors (abstract only). In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, Seattle, WA, USA, March 8-11, 2017*, page 722, 2017.

[4] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.