Infusing Data Science Across Disciplines

Debzani Deb

Department of Computer Science
Winston-Salem State University
Winston-Salem, NC, USA
debd@wssu.edu

Russell M. Smith

Department of History, Politics &
Social Justice
Winston-Salem State University
Winston-Salem, NC, USA
smithrm@wssu.edu

Muztaba Fuad

Department of Computer Science
Winston-Salem State University
Winston-Salem, NC, USA
fuadmo@wssu.edu

ABSTRACT

In this poster, we describe our effort to develop, pilot, and evaluate a model for infusing data literacy into undergraduate curricula across a variety of disciplines using a modular approach. Our pilot implementation achieved reasonable success in attaining student learning outcomes, enhanced engagement, and interests.

KEYWORDS: Data Science; CS+X; Computer Science Education

ACM Reference format:

Debzani Deb, Russell Smith and Muztaba Fuad. 2019. Infusing Data Science Across Disciplines. In *Proceedings of 24th Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE'19)*. July 15–17, 2019, Aberdeen, Scotland UK ACM, New York, NY, USA, 1 page.. <https://doi.org/10.1145/3304221.3325579>

INTRODUCTION

Data Analytics is an essential skill for twenty-first century workforce and as a result the need to help all students acquire such skill has recently gained increased attention. CS+X is one such approach of integrating data literacy into discipline-specific undergraduate courses through faculty partnerships. However, in most such initiatives, X is typically a STEM or engineering field and there are very few approaches that focus on enhancing the skills of social science or humanities students. Also, almost all CS+X courses focus exclusively on teaching students to program and therefore has a chance of isolating data literacy from disciplinary content and propagating the conception that “Data Literacy = Programming”. This may prevent social science faculties from forming a partnership with CS faculties in order to persistently and meaningfully infuse data literacy into their curriculum.

In this poster, we describe our effort to develop, pilot, and evaluate a model for infusing data literacy into undergraduate curricula across a variety of disciplines using a modular approach that does not rely entirely on programming. Rather than developing new discipline specific data science flavored courses, our approach integrates data literacy into existing courses through offering a week-long context-sensitive module and associated hands-on

activities. We believe that the modular approach is rather straightforward and easier to adopt than designing a special topic course and adopting it in a densely-packed curricula. Also, by adopting the data science modules in discipline specific core courses, there is a greater chance to reach a broader audience of students. The proposed modules are designed to be short and autonomous for easier adoption and as a result, each of them spans two 75-minute classes and includes specific learning goals, lecture notes, and assessment resources. The assessment resources contain quiz questions, tutorials, input dataset, and hands-on projects etc. The modules are expected to be taught in a lecture-lab setting, where the basics of data science are covered on the context of the target discipline and students are introduced to the hands-on project along with the necessary tools and resources, that they will require to complete the project. The take home project contains multiple tasks and usually has two weeks of deadline.

During the Fall of 2018, one such module is designed and piloted to a general education course titled as “Environmental Justice and Sustainability” at Winston-Salem state university. The course is mostly attended by social science majors (N: 24, Male:11, Female:13) and examines critical perspectives on social justice and geography. The learning outcomes of the module are as follows: 1) to understand how advances in technology enable the field of data science, 2) to be familiar with the sources of data relevant to the discipline, 3) to analyze a data set using spreadsheet and pivot table, and 4) to convey meaningful insights from a data analysis through visualizations. The take home project requires the students to explore the census and environmental data of Forsyth county (available at <https://www.epa.gov>), to use their spreadsheet and pivot table manipulation skills to answer ten questions such as “What is the average cancer risk for the East vs. West Forsyth?”, “What is the average respiratory hazard index for the older part of the city (more than 50% houses are built before 1960)?” and to provide necessary visualizations. Additionally, students are asked to complete a short six questions survey with some interesting results. While students overwhelmingly found data science to be interesting and could see the value of it for societal good, they remain mostly neutral about their desire to explore more of it in their future courses. The module and associated resources are available at <http://ibigcloud.altl.org>.

ACKNOWLEDGEMENT

This research was supported by NSF Award # 1600864.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. ITiCSE '19, July 15–17, 2019, Aberdeen, Scotland UK

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6301-3/19/07. <https://doi.org/10.1145/3304221.3325579>