Data Science for All: A Tale of Two Cities

Lillian Cassel

Villanova University

Department of Computing Sciences

Villanova PA 19085 USA

1 610 519 7341

Lillian.Cassel@villanova.edu

Don Goelman

Villanova University

Department of Computing Sciences

Villanova PA 19085 USA

1 610 519 7310

Don.Goelman@villanova.edu

Michael Posner

Villanova University

Dept. of Mathematics & Statistics

Villanova PA 19085 USA

1 610 519 3016

Michael.Posner@villanova.edu

Darina Dicheva

Department of Computer Science

Winston Salem State University

Winston Salem, NC USA

1 336 750-2484

dichevad@wssu.edu

Christo Dichev

Department of Computer Science

Winston Salem State University

Winston Salem, NC USA

1 336 750-2477

dichevc@wssu.edu

ABSTRACT

In this poster the authors report on experiences in teaching an introductory course in Data Science at two different institutions. Their approaches were informed by the aims of their NSF-funded project: to provide insight on learning goals, central data science topics, content modules, and a framework for implementing a flipped classroom approach to introduce data science to students with various technical backgrounds. The authors, investigators on the grant mentioned above, are a collaborative team of computer scientists and a statistician working to create flipped material for an introductory data science class. After ITiCSE the materials described in the poster will continue to be available in Ensemble, at <http://computingportal.org/datascienceflipped>

Keywords

Data Science, flipped classroom, computer science education

INTRODUCTION

Data Science programs of various granularities are emerging in many areas and are related to many disciplines, including any field dealing with massive amounts of data. This project looks at core concepts to provide resources for faculty teaching computing and statistics topics in the context of data science. The poster will review an overview of topics from both the computer science and statistics perspective, along with consideration of requirements related to domains. Its emphasis will be a description of, reflection on, and lessons learned from our experiences in delivering such a course at two institutions. Since we committed to producing learning materials, we chose to use the flipped classroom approach in the design of the materials. We believe that lack of suitable materials severely constrains the adoption of the flipped classroom approach, and we set out to address that in the context of the data science modules we produce.

Permission to make digital or hard copies of all or part 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'17, July 03-05, 2017, Bologna, Italy

© 2017 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-4704-4/17/07.

<http://dx.doi.org/10.1145/3059009.3073000>

Confirmation of the significance and relevance of this project was evident in well-attended (indeed overflowing) Birds of a Feather sessions at the SIGCSE Symposia in 2016 and 2017.

A mailing list of people interested in data science education now resides on a Villanova server. People who visit the poster will have an opportunity to join the mailing list and to use it for communication with the ever growing group.

The poster will include details on the pilot offerings by the authors, together with their reflection and recommendations. They will be described along various axes: the differences between the two institutions (Villanova University and Winston Salem State University); the relative roles of computing and statistics; student backgrounds; and the degree to which the flipped format was used.

A significant number of programs in data science now exist, and the poster will consequently include as well a representative sample of the program types, together with comparisons with our own experiences. There are several projects currently exploring curriculum issues related to data science. One of these is sponsored by the United States National Academies. Another project brings together a number of societies with interest in data science, including the Association for Computing Machinery (ACM), the American Statistical Association (ASA), the Association for Information Systems, The Association for Computers and the Humanities, The Business Higher Education Forum, the iCaucus, INFORMS, IEEE-CS, and others. Information about the meetings and conclusions of these groups are included in the poster.

ACKNOWLEDGEMENTS

This work is supported by NSF-DUE IUSE grants 1432256 and 1432438. In addition, we gratefully acknowledge the support of the graduate assistants who have worked on the project, including Dheeraj Kumar Reddy Kurla, Sai Vishwa Teja Mitta, Yamini Praveena Tella and LeelaRajesh Sayana.