Proc. 14th Int. Conf. on Global Research and Education, Inter-Academia 2015
JJAP Conf. Proc. **4** (2016) 011617

Innovative Way of Offering Master's Program on Data Analytics with Minimal Resources

Reneta P. Barneva[1], Valentin E. Brimkov[2], Joaquin O. Carbonara[2], John Favata[3], Barbara Sherman[3], and Kamen Kanev[4]

[1]*Department of Applied Professional Studies, The State University of New York at Fredonia, NY 14063, USA*
[2]*Department of Mathematics, SUNY Buffalo State, Buffalo, NY 14222, USA*
[3]*Department of Computer Information Systems, SUNY Buffalo State, Buffalo, NY 14222, USA*
[4]*Shizuoka University, Hamamatsu, Japan*

E-mail: barneva@fredonia.edu

(Received September 23, 2015)

Data Analytics is an area of high demand and the shortage of data scientists is becoming a serious constraint in some sectors. At the same time, universities often face budgetary restrictions, which do not allow them to hire faculty to establish the necessary new programs. In this paper we consider an innovative way of starting a new program with minimal resources considering a case-study of multi-campus, multi-department program on Data Analytics offered by SUNY Buffalo State and SUNY Fredonia. Each of the universities will offer some of the courses in synchronous or asynchronous distance education settings. In addition, some of the lecturers will come from various industrial sectors and students will be exposed to real-world problems. This makes the program very flexible and ensures invaluable practical experience which improves the marketability of the graduates. Combining the efforts of the two universities saves valuable resources and keeps the cost of instruction low. Shizuoka University, Japan, which does not have a program on data analytics, is studying the possibilities for sending students and engaging faculty members in the program.

1. Introduction

Data analytics is a novel discipline with applications in diverse fields: from health care, to engineering, to agriculture, to management, to marketing, to sport management, to sciences and to government.

Data scientist is considered a new profession requiring not only mathematical and statistical knowledge, but also technical skills and the ability to communicate the findings in terms of the specific domain. According to IBM "A data scientist will most likely explore and examine data from multiple disparate sources. The data scientist will sift through all incoming data with the goal of discovering a previously hidden insight, which in turn can provide a competitive advantage or address a pressing business problem. A data scientist does not simply collect and report on data, but also looks at it from many angles, determines what it means, then recommends ways to apply the data" [1]. For concrete details of the challenges, day-by-day work, and the necessary skills required by a data scientist, the reader is referred to the book *Data Scientists at Work* [2], consisting of 16 interviews with leading data scientists working for Facebook, LinkedIn, The New York Times, Quid, and others.

A pioneer article entitled "Data Scientist: the Sexiest Job of 21 Century" was published in Harvard Business Review of October 2012. It states: "A new role is fast gaining prominence in

organizations: that of the data scientist" [3]. The article continues: "data scientist is a high-ranking professional with the training and curiosity to make discoveries in the world of big data. […] On this front, demand [for data scientists] has raced ahead of supply. Indeed, the shortage of data scientists is becoming a serious constraint in some sectors." Further the article states: "No university programs have yet been designed to churn out data scientists."

In regards to the availability of university programs to train data scientists, a report at the conference "Training Students to Extract Value from Big Data" [4] organized by the National Academy of Sciences in April 2014 (just two years after the article mentioned above), listed a total of 30 programs nationwide. This number is very small compared to the demand for data scientists in industry and government. According to McKinsey report, "by 2018, the United States will face a shortage of up to 190,000 data scientists with advanced training in statistics and machine learning as well as 1.5 million managers and analysts with enough proficiency in statistics to use big data effectively" [5].

Gartner report [6] places this field in perspective expecting data analytics to emerge as a mature and well established academic and professional discipline. The jobs in data analytics are also ranked very high. For example, Tony Lee ranks this profession as the 6th highest ranking job with an average income of US$124,149.00 [7].

While there is a definite need for educating data scientists, in the last decade there exists a trend of decreasing the resources in the Academia, which obstructs starting new programs. In particular, if a new program is launched, it is expected to pass still at the very beginning an enrollment threshold, below which the classes at the respective university are cancelled. In order to get the program approved, the proposers should demonstrate that it has this potential. Yet, rarely new faculty are hired, before a program shows in practice that it attracts enough students. Thus, the first couple of years are critical for the participating faculty since they shall not only develop new courses in addition to their regular obligations, but take an active part in the program marketing and student recruitment activities as well.

Below we describe a model in which two universities offer jointly a program on data analytics. The main contribution of this model is proposing an innovative way of starting a new university program with very low initial resources. The approach is based on:

- Involving two or more institutions. Thus the resources and the efforts are spread among them and no hiring of new faculty is required before the program starts.
- Offering all of the courses through synchronous or asynchronous distance learning so that students from wider geographic area can attend.
- In view of the interdisciplinary character of the program, involving faculty from several departments.
- Including an internship providing real-life experience. In addition, inviting part-time instructors and mentors from industry.
- Massive use of technology and high flexibility.

The paper is organized as follows: In Section 2, we discuss details of the innovative model. In Section 3 the program structure is given. Finally, in Section 4 we conclude with some closing remarks and outline future plans.

2. Model of Offering the Program

The State University of New York (SUNY) is one of the largest university systems in the world, encompassing 64 campuses, covering New York State in a dense network. SUNY Buffalo State and SUNY Fredonia are about 50 miles (80 km) apart. While collaboration among faculty is established, it is difficult for the students to commute due to the lack of public transportation. This is especially true in the winter weather conditions, which may start as early as October and last until April. Both campuses possess talented educators in mathematics, statistics, computer science, and information

systems – the areas necessary for starting an interdisciplinary program in data analytics – but at the same time, each of them does not have the resources to offer a new program on its own.

Professional Science Masters (PSM) is a relatively new kind of graduate degree that combines a specific application area with business skills to ensure the graduates will do well in the "real" world. PSM programs have advisory boards from the industry that work closely with faculty to make sure the curriculum is relevant to economic realities, and help provide students with internships and eventually with jobs [8]. Some of the employers may also teach courses in the program as part-time instructors and an important part of the studies is based on applied learning.

Buffalo State is one of the pioneers in PSM, implementing two PSM programs in the last few years. Faculty from Mathematics and Computer Information Systems Departments have designed a new PSM degree program in Data Analytics Across the Disciplines (DAAD), which will prepare students with a background in the STEM (Science, Technology, Engineering, and Mathematics) disciplines to join the workforce in Information Technology (IT), in particular as Data Scientist/Analyst (DS/DA). Faculty with expertise in statistics and business have joined the interdisciplinary team to provide the students with the necessary skills in these two areas.

Fredonia has a very successful experience in a two campus collaborative delivery of courses, offering a course on Programming for Embedded Microcontrollers through simultaneous distance learning to students in two SUNY campuses more than 200 miles (300 km) apart. The course was offered in a synchronous virtual classroom based on web-conferencing systems allowing interaction in real time with multiple users. Such rooms are commonly used to conduct meetings or interviews and most of the universities in the developed countries are equipped with them. Web cameras, microphones and speakers allow the users to see and hear each other. Other features as text chat, application sharing, breakout rooms, and interactive whiteboard are often also available. Although the course on Embedded Microcontrollers includes programming and assembling hardware parts, in which the help of the instructor is indispensable, he was able to provide it in a timely manner as if he were physically among the students.

Thus, at the early stage of PSM-DAAD program development, the educators from Fredonia and Buffalo State joined resources and expertise to provide adequate course delivery. The courses in the program will be run in synchronous virtual classrooms. Lectures and activities can be recorded for dissemination. Asynchronous online versions of the courses will also be developed as needed. Both campuses will advertise the program and the course offering will be held at the two universities, thus leveraging the key resources.

3. Program Structure

The program outline is given in Table I below:

Table I. PSM Data Analytics Across Disciplines to be offered by SUNY Buffalo State and Fredonia.

PSM-DAAD
Semester 1
1) (Core) DS 500 Introduction (Overview) of Data Science - overview of the field - topics covered at a somewhat superficial level but wide spectrum
2) (Core) DS 510 Review (Overview) of Mathematics for Data Science - Discrete math, Calculus I, II, Elements of Linear Algebra
3) (Core) DS 520 Elements of Computer Science (CIS) and Programming for Data Science - Algorithms, Languages, Operating Systems, Networking, and others

PSM-DAAD (cont'd)

Semester 2

4) (Core) DS 530 Introduction to Databases for Data Science
- SQL, relational algebra, databases
- NoSQL, Flat Tables, etc.

5) (Core) DS 540 Intro to Statistics for Data Science
- intro to probability
- classical statistics
- analytics, inference, predictive analytics
- elements of Machine Learning

6) (Core) DS 550 Business, Entrepreneur and Leadership Skills

Semester 3

7) (Analytics track) DS 600 Machine Learning Techniques for Data Science
- regression techniques
- classification techniques
- Bayesian techniques, etc.

8) (Analytics track) DS 610 Big Data Management and Distributed Database Systems
- large data warehousing
- "Cloud" computing
- Hadoop and similar distributed/parallel systems
- current data tools overview, etc.

9) (Analytics track) DS 620 Visualization and Presentation Techniques
- n-dimensional plotting/representation
- presentation strategies/perception, etc.

Semester 4 (all Required for MS degree)

10) DS 630 Elective
- some topic not in the above list, advanced math, data science tools, etc.

11) DS 640 Major Project (supervised)
- significant project in some application area of interest to student

12) DS 650 Internship
- with a business partner

This structure serves as a template for future development of the curriculum. Depending on the needs of the student and institution, an abbreviated program version can be offered. For example, some MS programs are limited to 30 credits. The program outlined in Table I is designed for 36 credits but can be modified. In Semester 4, by eliminating DS 630 Elective and folding parts of DS 640 Major Project into DS 650 Internship, a 30 credit program can be offered. In addition, a Data Science Certificate can be offered by requiring students to complete the core courses offered in the first two semesters (18 credits).

Flexibility is another aspect of this program. Its ability to offer tracks enhances the value of the program and allows faculty from various departments and sister institutions to join in the curriculum. Semester 3 serves as the implementation point for tracks. So far, we are designing the following tracks: Scientific Data Analytics, Educational Data and Learning Analytics, Business

Analytics, and Geographical Information Systems (GIS) Analytics. Other tracks are under consideration. To minimize impact on resources, these tracks utilize existing departmental courses, but refocus them into the PSM theme.

4. Program Perspectives

4.1 Admissions

We assume that students enter the program with various levels of mathematical and computational skills. This program will have admission requirements, including a STEM undergraduate degree and a significant Graduate Record Examination (GRE) score. However, waivers may be granted for promising and ambitious candidates. Significant deficiencies by otherwise strong candidates will be ameliorated by directing students to other existing institutional courses or remedial programs. The task of student advisement throughout the program will be done by program faculty.

4.2 First Year

The first five core courses introduce students to the necessary background for DS/DA professionals. The main goal of DS 500 Introduction of Data Science is to give a broad perspective on current activities in data science. Students will be introduced to a wide variety of contemporary "big" data problems selected from a number of academic, business, health, and industry fields. Expert guest-lectures will be solicited from workers in the field. It is hoped that such exposure will help students develop their interest for a particular application area. DS 510 Review of Mathematics for Data Science gives students the necessary mathematical background for DS/DA workers. Since data science touches many areas of mathematics, we have narrowed this course to a useful subset of topics including elements of probability, calculus, optimization, linear algebra, discrete mathematics and numerical methods. DS 520 Elements of Computer Science (CIS) and Programming for Data Science provides a background in selected topics of Computer Science and Information Technology. Students will obtain skills in selected technical aspects of current operating systems, programming languages, introduction to algorithms, data processing and storage, and communications. DS 530 Introduction to Databases for Data Science introduces students to contemporary database technologies, including extensions to distributed systems and non-traditional approaches. DS 540 Introduction to Statistics for Data Science introduces students to the necessary statistical theory and skills needed for professionals. Topics will include computational techniques for large scale problems. DS 550 Business, Entrepreneur and Leadership Skills introduces students to the world of business as most STEM students generally have had very little formal exposure to business. There is a general consensus that business skills are part of a well-rounded DS/DA professional and selected business topics will be introduced here.

4.3 Second Year

The second year is where important professional skills are developed. In the first semester, track courses are offered. Students are fully immersed in their track and work with faculty on data problems in the track. It is possible that selected research topics can be introduced but this is not the main focus of the PSM program. In the second semester, students polish their skills by engaging in an elective, project and internship. The elective course can come from any available academic master's program assuming course prerequisites are met. It is anticipated that students will take their elective course from the department that offers their track, since prerequisites will likely be met. The major project, which is faculty guided, challenges student creativity and technical skills. The internship with an external partner polishes technical and business skills of the student and ties the program together. In addition, possibilities for continuing employment can be mutually explored by students and partners.

5. Conclusion

The Buffalo State and Fredonia administrations are very supportive of the program. A mini-conference was held to announce and popularize the program. It was attended by over 50 administrators, faculty and students from three SUNY campuses as well as by representatives of the industry, financial sector, and local government. The program is initially scheduled to run at Buffalo State and Fredonia under the Multidisciplinary MS degree schema at first. The program will start with courses that have already being taught at Buffalo State and Fredonia, thus providing students with certain necessary prerequisites. Meanwhile, in AY 2015-2016, the program will be submitted for formal approval. The synchronous and asynchronous online courses will eventually be available to Open SUNY.

Possibilities for admitting international students to the program have also being investigated. Students from Shizuoka University, for example, could attend classes on-campus at Buffalo State and/or Fredonia while on a visit and continue with remote classes after their return to Japan.

Acknowledgments

This work was supported in part by funding for a Cooperative Research Project at Research Institute of Electronics, Shizuoka University and KAKENHI Grant 25560109. The authors thank the anonymous referees for the useful remarks, which helped improve the paper.

References

[1] What is a data scientist? IBM, http://www-01.ibm.com/software/data/infosphere/data-scientist/ (Retrieved December 2015).
[2] S. Gutierrez, Data scientists at work, Apress (December 2014).
[3] T.H. Davenport and D.J. Patil, Data scientist: the sexiest job of 21th century, Harvard Business Review **90**(10), pp. 70-76 (October 2012).
[4] M. Mellody, Training students to extract value from big data: summary of a workshop, National Research Council (2014).
[5] S. Lund et al., Game changers: Five opportunities for US growth and renewal, McKinsey Global Institute Report (July 2013), http://www.mckinsey.com/insights/americas/us_game_changers, (Retrieved December 2015).
[6] F. Buytendijk, Hype cycle for big data, Gartner, G00261655 (August 2014).
[7] T. Lee, Jobs rated report 2015: ranking the top 200 jobs, http://www.careercast.com/jobs-rated/jobs-rated-report-2015-ranking-top-200-jobs, (Retrieved July 2015).
[8] NPSMA: National Professional Science Master's Association, http://npsma.org, (Retrieved July 2015).