REPORT ON NSF-FUNDED GRADUATE TRAINEESHIP IN DATA SCIENCE TECHNOLOGIES AND APPLICATIONS

B. Furht, R. Tappen, T. Khoshgoftaar, H. Kalva, O. Marques, X. Zhu, J. Jang, M.T. Jan, N. Romance, and M. Pinto

Florida Atlantic University (UNITED STATES)

Abstract

Data science and analytics are emerging transdisciplinary areas comprising computing, statistics, and various application domains including medicine, nursing, industry, and business applications among others. A significant shortcoming of the current graduate curriculum in the U.S. is that scientists and engineers are well trained in their own areas of specialty but lack the integrative knowledge needed for new scientific discoveries and industry applications made possible by data science and analytics. To address these shortcomings, we proposed to NSF a new model of convergence education through experimental learning consisting of testbeds, bootcamps, and case studies. Our main goal is to create a novel graduate curriculum with intensive convergence activities, which also include transdisciplinary research projects. In 2020, we received $2.4 million in funding from NSF to establish and launch such a graduate program and train future data scientists from various disciplines. In this paper we describe the program and report results based on two cohorts of graduate students who successfully completed the program and received their master's degrees in data science and analytics.

Keywords: Data science, data analytics, transdisciplinary curriculum, normalization courses.

1 INTRODUCTION

Data science is often described as the intersection between and among mathematics, statistics, programming skills, business intelligence, and domain expertise. While students pursuing graduate degrees in science and engineering typically develop skills in these areas, they often lack practical training on how to apply them to data science research and industrial tasks, particularly where traditional data processing and analysis techniques are not viable due to increased volume, variety, and complexity of the data. A recent article discusses data scientist challenges from a business perspective [1]. The author claims that the best data scientists are not just statisticians or machine learning experts; they are also authorities in the field or business where they are applying those skills. Effective data scientists need to be able to work in interdisciplinary teams and to use data visualization and communication skills to communicate their findings to individuals not trained in data science. It is apparent that current graduate training in data science does not sufficiently prepare students. To address these shortcomings, we proposed a new model of convergence education through experimental learning consisting of testbeds, boot-camps, and case studies.

Our main goal was to create a novel graduate curriculum with intensive convergence activities, which also included transdisciplinary research projects. We have applied a transdisciplinary approach, which is radically distinct from multidisciplinary and interdisciplinary approaches [2]. A transdisciplinary approach, which we implemented in organizing our graduate curriculum, assumes dissolving the boundaries between conventional disciplines and organizing teaching and learning around the construction of meaning in the context of real-world problems [3]. The intent is to train students to collaborate with others inside and beyond the academy and to analyze and solve real-time world problems.

The main training elements of our transdisciplinary curriculum included the development of normalization courses and the creation of different testbeds for various application domains. Normalization courses were used to address various backgrounds of students entering the program. Our transdisciplinary and convergent research themes were focused on three data science and analytics areas: (i) medical and healthcare applications, (ii) industry applications, and (iii) data science technologies. To address these, we created a transdisciplinary curriculum for graduate students in data science and analytics, where each course was developed by at least two faculty members from two different disciplines. To integrate research and training, we used a Data Science and AI Laboratory to

develop multiple testbeds for different application domains. Each testbed included a computer platform, software tools, and set of learning modules. We provided a detailed evaluation and assessment plan including a logic model to anticipate outcomes and perform formative evaluation. A total of 14 trainees funded by the NSF completed the program in the last 2 years, received graduate degrees, and published research papers and patents based on research conducted as the part of the program.

2 VISION AND GOALS

While data science technologies and applications have evolved significantly over the last several years, it is clear that the current graduate training in data science does not sufficiently prepare students for the future challenges as researchers and practitioners in data science and its applications. The drawbacks of the current graduate training in data science can be summarized as: (1) the graduate curriculum is not adequate, and (2) students do not experience sufficient convergence activities as part of their training. Convergence education through experimental learning is critical in training future data scientists. The goals of the proposed NRT project are described next.

Goal 1. New Transdisciplinary Graduate Curriculum

We developed a new graduate curriculum based on transdisciplinary courses and research projects and used extensive convergence activities. The new program was developed for both Master's and PhD students from various departments and colleges including Computer Science, Computer Engineering, Electrical Engineering, Civil Engineering, Nursing, Science, and Medicine. The Master's program in Data Science takes 18 months, while the PhD program takes 3 to 4 years, which is typical timing for graduate programs. The program includes normalization courses, boot-camps, in-depth elective courses, testbeds, and professional workshops. The detailed curriculum is described in the section on "Education and Training." The result of this approach produced graduates with technical depth and transdisciplinary understanding of data science technologies and applications.

Goal 2. Transdisciplinary Research Projects

Students in the NRT program are involved in interdisciplinary research projects. Testbeds are developed in our newly created AI-DA laboratory. Research projects were formulated jointly with industry partners who are members of our NSF I/UCRC CAKE, so students worked on industry-focused problems. The desired result of this goal was to advance data science and technologies and address societal grand challenges in this area.

Goal 3. Initiate new activities to recruit and retain women and other underrepresented groups.

We developed new activities to recruit students for the NRT program with a focus on women and underrepresented groups. The U.S. Department of Education, Office of Post Secondary Education recently informed us that FAU is designated as an eligible institution under Titles III and V. This means that FAU is a Developing Hispanic-Serving Institution (HSI) and is eligible for Promoting Postbaccalaureate Opportunities for Hispanic Americans (PPOHA).

Florida Atlantic University was ideally positioned to implement the proposed new model of education and research based on transdisciplinary courses and research projects and applying convergence activities. Here are the strategic institutional factors supporting the project:

1. Strong Administrative Support from the University President and College Deans: The FAU President pledged strong FAU support for this project, and he is committed to laboratory support including a $1 million gift.
2. Transdisciplinary Curriculum: The program and TD curriculum include bold and novel academic approaches including normalization courses, boot-camps, in-depth elective courses, testbeds, and professional workshops.
3. Industry Collaboration: FAU is the home of the NSF I/UCRC CAKE under directorship of PI Furht, which is in its tenth year of operation conducting applied research in the areas of data science technologies and applications. Industry members are large and small companies that are involved in products, services, and applications in the field of data science, which provides a great opportunity for NRT students to work with industry collaborators. In addition, five years ago FAU created Tech Runway, which provides an opportunity for faculty and students to start-up companies based on their research. Therefore, some research projects, as part of the NRT proposal, will be conducted in collaboration with FAU Tech Runway, which will provide a great opportunity for NRT students to work on real world problems and learn about entrepreneurship.

4. Strong Research Focus on Data Science and Analytics: FAU researchers have already made strong research contributions as well as academic results in data science technologies and applications. FAU is the only university in Florida that was recently selected by the Florida Board of Governors as a University of Distinction in Artificial Intelligence and Big Data Analytics. This area is also a platform of FAU's strategic plan approved by the Board of Trustees and Board of Governors.
5. FAU is a Designated URM Institution: FAU has the unique advantage of having a Hispanic Serving Institution designated by US Department of Education. We have leveraged this advantage in recruiting URM minorities and women and creating NRT data science teams of students which are diverse ethically and by gender.

3 TECHNICAL PRINCIPLES AND COURSES

The transdisciplinary education in data science and analytics cannot be accomplished by simply collecting existing courses from different departments into a new degree program. Therefore, we proposed the transdisciplinary approach in which students must learn how the different disciplines interact and relate to each other. Our solution applied innovative approaches that included: (a) developing normalization courses, (b) developing in-depth transdisciplinary courses, and (c) creating different testbeds from different application domains.

Technical topics were incorporated in two types of courses: normalization courses and in-depth elective courses. Normalization courses served to provide students from different backgrounds with the basics of data science and analytics and their applications. They also provided a strong foundation in data science skills including statistics, probability, mathematics, data mining, machine learning, and predictive analytics. The testbed approach provided students with hands-on experience ~~in~~ across several critical data science applications. Table 1 shows the proposed transdisciplinary curriculum consisting of normalization courses and in-depth elective courses.

Table 1. Transdisciplinary Curriculum in Data Science Technologies and Applications.

Normalization Courses	In-Depth Elective Courses
• Introduction to Data Science • Data Mining and Machine Learning • Quantitative Methods • Biomedical Data and Informatics • Introduction to Business Analysis & Big Data Methods in Human Subject Research	• Social Networks and Big Data Analytics • Web Mining • Computational Advertising and Real-Time Data Analysis • Applied Statistical Methods • Regression Analytics • Deep Learning Techniques • Advanced Data Mining and Machine Learning • Cyber Security Measurements and Data Analysis • Data Analysis for Managers • Advanced Business Analytics • Spatial Data Analysis • Artificial Intelligence • Statistical Analysis for Data Science

Table 2 shows a typical 3-semester schedule for Master's students (it includes the summer semester), and Table 3 shows a program schedule for PhD students. Master's students will take normalization courses in the first two semesters and will take in-depth courses as needed. The Master's program at FAU requires a total of 10 courses or 30 credits, so students will complete it in three regular semesters, which is the same timeline ~~as~~ for all ~~other~~ Master programs.

Table 2. Typical Course Schedule for Master's Students (Total 10 courses, 30 credits).

Semester 1	Semester 2	Summer Semester	Semester 3
3 normalization courses Bootcamp sessions	2 normalization courses + 1 in-depth elective course or 3-in-depth elective courses	1 in-depth elective course	3 in-depth elective courses

PhD students typically take between 1-3 normalization courses, depending on their background, and the remaining in-depth elective courses. The total course requirements for PhD students are 6 courses (18 credits) plus Dissertation research (33 credits), which totals 51 credits. Typical PhD study at FAU is 6-8 regular semesters or 3-4 years, so in the remaining 4-6 semesters PhD students will complete the qualifying exam, conduct their transdisciplinary research on the projects proposed in this program, and work on their professional development and communication skills.

Table 3. Typical Course Schedule for PhD Students (Total 6 courses,18 credits), and Dissertation Research (33 credits).

Semester 1	Semester 2	Semester 3+
1-3 normalization courses Up to 2 in-depth elective courses Bootcamp sessions	3 in-depth elective courses	Qualifying exam Dissertation research Professional development Communication training

Testbeds, bootcamp, and case studies

We developed multiple testbeds from different application domains. The focus has been on the following application modules: medical and healthcare, industry applications, and data science technologies including AI. Each testbed included (a) a computer platform, (b) software tools, and (c) a set of learning modules. The learning modules were designed ~~in such a way~~ to teach specific concepts in each course. The course employed hands-on projects~~,~~ and team-based educational methods. Application themes and the testbeds are based on the participating faculty members' active and funded transdisciplinary research projects in data science and analytics.

Bootcamp sessions are organized every semester for new students to get them familiar with the platforms for various testbeds. The bootcamp consisted of student teams from different departments working with 2 faculty members from different disciplines.

In summary, our transdisciplinary graduate program in data science technologies and applications has the following characteristics:

- The program integrates training and research while not extending the time to complete the degree.
- It includes a new transdisciplinary approach beyond disciplinary perspectives.
- It involves women and students from underrepresented groups into STEM fields.
- It provides convergence education through experimental learning.

4 MAJOR RESEARCH EFFORT

The NRT is motivated to integrate education and research and train graduate students in the six research Big Ideas recommended by NSF. Specifically, we leveraged research conducted in our NSF Industry/University Cooperative Research Center for Advanced Knowledge Enablement (CAKE) at FAU [4]. The CAKE Center was established in 2009, and in a ten-year period since its founding, had 45 industry members with total funding of about $8 million. The Center successfully completed 55 applied industry projects. The research theme of the Center is in the field of advanced knowledge enablement with a focus on areas of big data analytics, multimedia and data mining, video and multimedia processing, and cloud computing systems, sensors, and networks. Research projects are interdisciplinary in their nature with a focus on applications in medical and healthcare, environmental, and mobile applications, along with public~~,~~ and industrial applications. These are identical to the synergistic activities needed to complete the applied research projects that we conducted as part of research in the NRT project related to Harnessing the Data Revolution (HDR).

Figure 1 illustrates research areas and related vertical applications. This section presents ongoing and envisioned projects in the following three data science and analytics areas: (1) Medical and Healthcare, (2) Industry Applications, and (3) Data Science and AI Technologies. Within each of these areas, challenges and specific project examples are presented. Most of the presented projects are proposed by CAKE industry members and are funded or will be funded by industry memberships. Therefore, practically all projects involve collaboration with application domain experts and industry partners.

Students' interaction with these experts and industry researchers provides another key component of the NRT project, which is convergence training.

NRT professors have already contributed to the areas of NRT research with many contributions [5-8]. Medical and healthcare projects, which we investigated as part of the NRT included (i) Analyzing and detecting cognitive changes in older drivers using in-vehicle sensors and data analytics, (ii) Melanoma and brain tumor detection using data analytics techniques, and (iii) Medicare fraud detection using a big data set. In industry applications, we investigated: (i) The development of machine learning algorithms for big data on an industry platform referred to HPCC/ECL, (ii) Application of machine learning algorithms in automotive industry, and (iii) Developing new video coding algorithms using big data analytics.

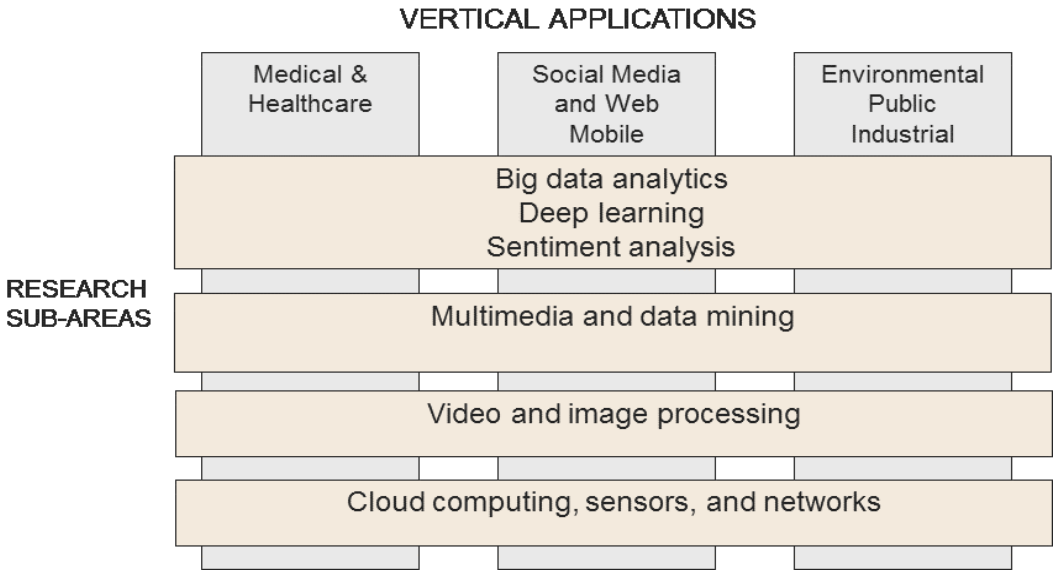

Figure 1. Research areas and related applications conducted within the NRT project.

5 ORGANIZATION AND MANAGEMENT

Management organization of the NRT program is shown in Figure 2. The NRT program is directed by the Principal Investigator of the project. The Project Coordinator was hired and reports to the Director. The Project Coordinator is responsible for overseeing four committees and internal evaluations. The PI worked with all committees to provide the direction and vision for the NRT program. Jointly with the Project Coordinator, PI is also responsible for the Outreach Program.

The Evaluation and Assessment Committee has been coordinated by the Project Coordinator and the evaluation is carried out by an external evaluator from Rockman et al.

The Recruiting and Student Development Committee consists of two subcommittees, which are responsible for recruiting a diverse group of students from different departments and providing professional development for these students.

The Curriculum Committee is responsible for the development of the program curriculum, and related normalization and in-depth elective courses.

The Research Committee provides guidance for research topics for NRT students. It collaborates with industry partners relating to these projects and coordinates the development of testbeds.

The External Advisory Panel consists of representatives from industry corporations who provide advice and industry perspectives of the NRT program and its progress.

Recruitment, Mentoring, and Retention. The proposed NRT program is designed to include students from five colleges, and a dozen departments from these colleges. The colleges whose faculty and students participate in the NRT program include (i) Engineering and Computer Science, (ii) Medicine, (iii) Science, (iv) Nursing, and (v) Business College.

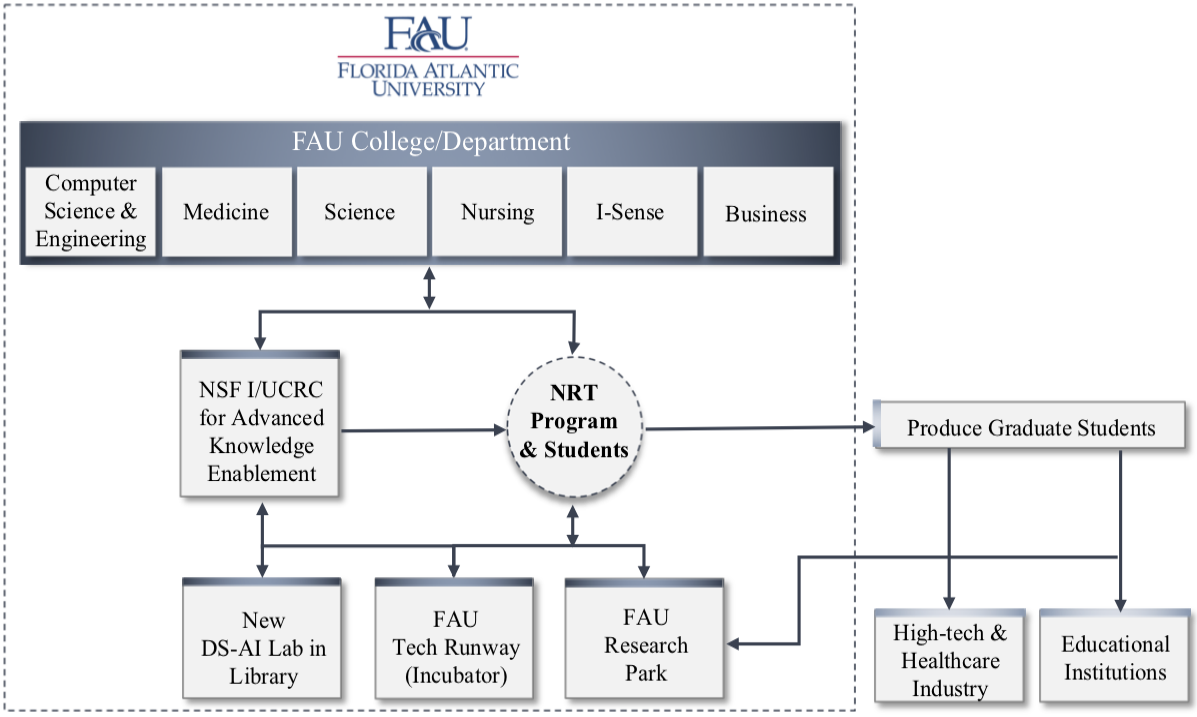

Figure 2. Management organization of the NRT program.

6 PERFORMANCE ASSESSMENT AND PROJECT EVALUATION

First two years Rockman et al (REA) conducted the external evaluation of the *NRT-HDR: A Graduate Traineeship in Data Science and Analytics* program. The company has extensive experience evaluating graduate training programs, often for NSF-funded projects.

In order to inform the design of the transdisciplinary data science and analytics curriculum and program supports, REA, in its initial years, focused its efforts towards gathering information from participating graduate students and faculty on the most and least effective program features and activities. Their feedback was used to help guide the project team in making data-driven decisions around possible program modifications that would best maximize positive faculty and student experiences and outcomes. In the final years of the program, REA will turn its attention towards the impact~~s~~ of the program on NRT participants in terms of changes in their interdisciplinary understanding of data science and its applications in industry, ~~and~~ in the medical and healthcare fields, and their confidence ~~around~~ and actual participation in transdisciplinary research. During this summative phase, REA also explored possible impacts on faculty participants regarding changes in their mindset towards using transdisciplinary approaches for graduate student training.

Although both the formative and summative phases of the evaluation have different foci, the proposed methods to collect data across the project are similar. REA gathered data via multiple methods:

Pre- and post-surveys of participating graduate students to track changes in their understanding of data science and analytics, participation in transdisciplinary program activities, and confidence in their developing skills.

- Focus groups with graduate students to gather more in-depth information about their experiences within and the impacts of the program; and
- Interviews with faculty to obtain their perspectives on the program's activities and impacts – both on themselves and on participating graduate students.

The planned pre-surveys will be conducted within the first few weeks that each new cohort of students enters the program, while post-surveys will be administered annually in the Spring of each program year. Faculty interviews are scheduled at the beginning of the project to determine the extent to which program goals and expectations are aligned and annually in the Spring of each program year to provide

faculty with a chance to reflect on the progress made to-date as well as to highlight any planned or implemented program modifications that may affect project outcomes. Student focus groups will occur annually in the Spring of program years 2-4 to help identify program characteristics and practices that might be improved or changed to better serve program participants. These focus groups are paired with site visits that give the evaluation team a chance to observe any culminating program activities, as well as the opportunity to meet with the project team to reflect and plan in-person.

Survey results, interviews and focus groups are used as formative program evaluation measures to gather data on program characteristics and practices that can be improved or changed. Findings are shared with the project team annually via memos that are appended to their Annual Report to NSF, and also serve as discussion documents for decision-making. A final summative report is submitted to the project team in the last year of the project highlighting program outcomes and lessons learned. The detailed logic model diagram that includes inputs, activities, outputs, outcomes, and impact of the proposed NRT-HDR program is shown in Figure 3.

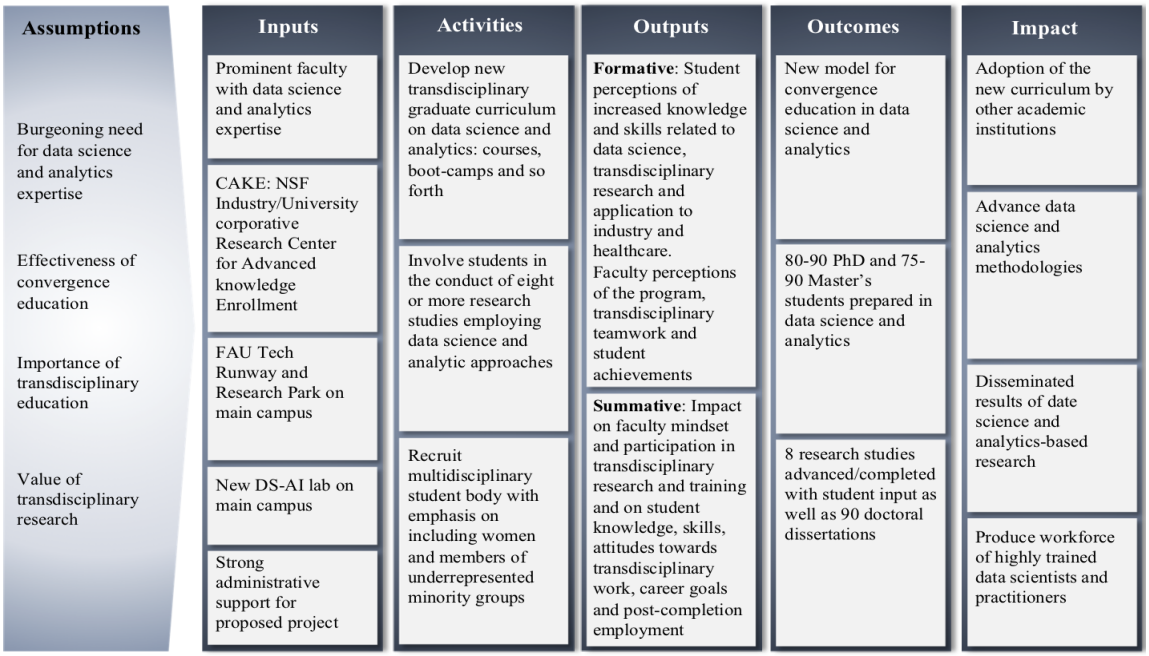

Figure 3. NRT Logic Model.

We implemented the logic model for the NRT Cohort 1 and 2 students. Starting with "Inputs" we completed the following: (1) We selected our faculty with data science and analytics experience. Besides Computer Science faculty, we also selected several faculty members from other disciplines who are already using and teaching data science technologies and applications. (2) We actively collaborated with NSF I/UCRC CAKE Center in providing research participation of the NRT students in CAKE industry projects. (3) We collected and analyzed various metrics, which are important for the focusses of the program including: (a) total number of students enrolled in the program, (b) number of women enrolled in the program, (c) number of students from underrepresented groups enrolled in the program,(d) number of projects completed, (d) number of papers published, (e) number of PhDs and Masters' students, and (f) number of faculty who participated in the program.

7 RESULTS

In this section, we present results of implementing the program for two cohorts of NRT trainees in the period 2021-2022 (Cohort 1) and 2022-2023 (Cohort 2) [9].

A total of 18 graduate students successfully completed the program (14 fully funded by NSF and 4 not NRT funded). In addition, in our NRT transdisciplinary courses, we had, on average, about 45 graduate students in each class.

The focus on research projects from Cohort 1 and 2 resulted in several published papers and patent applications that are authored by the NRT students and their advisors [10-16].

The outcomes of the program and graduates from Cohort 1 and 2 are presented in the NRT Website at: https://www.fau.edu/engineering/eecs/research/nrt/news/. Success stories were published on the Website as press releases. Here we describe several successful stories. Juan Merlos, an NRT student, has been the first inventor of the patent application presented at the MPEG standard committee meeting for the VCM standard in Geneve, Switzerland in July 2023 [16].

Nicholas Marques, an NRT student, participated in the research project whose goal was to determine if the machine learning methods are useful to predict opioid use disorder and investigated how the research papers trained and evaluated these models. Findings, published in the journal ***Computer Methods and Programs in Biomedicine*** [15], reveal that while results from the reviewed papers show machine learning models applied to opioid use disorder prediction may be useful, there are important ways to improve transparency and reproducibility of these models, which will ultimately enhance their use for research.

Yherzon Chura Ruiz, an~~d~~ NRT student, launched a company, aDubb's, whose mission is to empower end-consumers of media content to access automatically generated dubs in their language of choice. Th~~is~~ model is breaking the language barriers so that users have access to foreign language media with the help of proprietary AI models and systems. In August 2023, Therzon entered the Business Pitch Competition, where he received *2nd place.*

The major evaluation findings, reported to NSF [17], can be summarized as follows:

1. The NRT program is addressing a major industry need, namely, the preparation of graduates who have a developed skill set on the technical aspects of data science/analytics across a broader spectrum of application areas such as healthcare medicine, neuroscience, and business.
2. From the initial survey to the post survey, the NRT students continued to indicate that their long-term career goal favored employment in industry, with government employment ranking second and academia ranking third.
3. From the pre-to-post survey, the NRT students maintained high confidence in the program's ability to prepare them for their desired career trajectory.
4. The program excelled in terms of the quality of its academic program.
5. The survey results indicated that a few students were still interested in increasing their confidence in areas such as their research methodology, including developing compelling research proposals.

7.1 Broader Impact

Data science and analytics are an emerging transdisciplinary area comprising computing, statistics, and various application domains including medicine, nursing, industry and business applications, and others. Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is related to data mining and big data. Data science technologies and their applications are already transforming our world and economy. We describe several major impacts that we believe that our program will have.

(1) Impact on Data Science and Analytics Community

The proposed new model of convergence education through experimental learning consisting of testbeds, bootcamps, and case studies is designed to transform education for the future generation of engineers and data scientists. As part of the NRT project, we documented and described the process of developing the program and instructional structure necessary to create a successful DSTA program. Analyzing, documenting, and disseminating the process will contribute to the development of similar programs in other educational institutions.

(2) Impact on Broader Data Science Community

By integrating research and education and working on a variety of real-life research projects jointly with industry partners, we produced and disseminated various publications relating to transdisciplinary projects and reporting the obtained research results. This can impact the broader data science

community. In addition, we also shared with the data science community the curriculum development for a DSTA program, course slides, tests and homework, testbed material, lab assignments, exams, and students' reports.

(3) Impact on US Competitiveness by Creating Highly Trained Data Scientists and Practitioners

- Our program has already produced highly trained data scientists and practitioners both for US industry and educational institutions The demand for data science skills is growing exponentially, and hence people who possess strong data science skills are in rising demand. Job openings for the role of data science has skyrocketed by 600% since 2012. Data analytics is becoming mission-critical to more and more businesses, and one of the biggest challenges they face is recruiting data scientists. Allen Blue, co-founder of LinkedIn, reported that data science and machine learning related jobs represent five of the top 15 growing jobs in America today. In summary, US companies will have access to highly trained graduates in data science technologies and applications, which will help them to remain competitive in the global marketspace.
-

(4) Impact on Women, Minorities, and Underrepresented Groups

Our program has a great impact by reaching out to women, minorities, and underrepresented groups. FAU has a large Under-Represented Minority (URM) population, and it is ranked as the most diverse public university in Florida and second in the Southeast. As already indicated, FAU has been designated as a Hispanic Serving Institution. We have already developed innovative activities to recruit students from these groups, which resulted in a diverse group of the NRT students.

8 CONCLUSIONS

In summary, our NRT transdisciplinary program includes several mechanisms to provide formal training in transferable professional skills. One component of our program is the required convergence training through research and course-based project experiences. Data science and analytics applications are highly multidisciplinary and require such training. Our proposed convergence training is based on multidisciplinary teams, industry created testbeds, and real-world application-driven problems. Students work side-by-side with students and faculty from multiple disciplines, rotate through project management roles during the projects, and interact with industry by considering translational and entrepreneurship possibilities of the projects and associated research. Students are also required to present weekly reports to the group and periodically to industry partners. Feedback for teamwork, communication, and project management are given by project leaders, instructors, and team members. Instructors also provide such feedback with grades at the end of class-based projects. Industry partners evaluate the translational and entrepreneurship aspects of the work. These testbeds and related research projects are integrated with the courses, so that students are exposed to such training multiple times during their education.

The second component is the required bootcamps. The bootcamps have two goals: (1) technical introduction to the details of the testbeds, and (2) an introduction to soft skills. For this latter goal, the required bootcamp provides instruction on teamwork, project management, and entrepreneurship. For entrepreneurship, we made use of the collaboration with the incubator in the FAU Tech Runway.

We expect that the proposed program will drive graduate education in data science and analytics nationwide.

ACKNOWLEDGMENTS

This paper is produced with the support of the NSF award #2021585 entitled NRT-HDR: Graduate Traineeship in Data Science Technologies and Applications, 2020-2025, $2.4 million.

We thank Drs. Raquel Assis, Behnaz Ghoraani, Mehrdad Nojoumian, Dingding Weng, Adam TM Wyatt, and Saeed Rajput, all from Florida Atlantic University, for participating as instructors and mentors in the NRT program. We also thank the following external participants from academia and industry, who participated as guest speakers in the NRT program: Drs. Flavio Villanustre and David Jaramillo from LexisNexis, Dr. Alfredo Cateriano from JM Family Enterprises, Dr. Mihai Fonoage from Modernizing Medicine, Dr. Giti Javidi from University of South Florida, Dr. Anton Kos from University of Ljubljana,

Slovenia, Dr. Tessa Harland from Albany Medical College, New York, and Dr. Megan Coffee from Colombia University.

REFERENCES

[1] J. L, Bradford, “8 Real Challenges Data Scientists Face,” Forbes Magazine, September 6, 2018.

[2] O. Guvenen, “Transdisciplinary science methodology as a necessary condition in research and education,” Transdisciplinary Journal of Engineering & Science, Vo. 7, pp. 69-78, 2016.

[3] In B. Nicolescu (Ed.) Transdisciplinarity – Theory and Practice, Hampton Press, Cresskill, NJ, UL

[4] NSF Industry/University Center for Advanced Knowledge Enablement at FAU, http://cake.fau.edu.

[5] B. Furht and F. Villanustre, “Big Data Technologies and Applications,” Springer Science and Business Media, New York, 2016.

[6] M.M. Najafabadi, F. Villanustre, TM Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” Journal of Big Data, Vol 2., No.1. https://doi.org/10.1186/s40537-014-0007-7, 2015.

[7] V.M. Herrera, T.M. Khoshgoftaar, F. Villanustre, and B. Furht, “Random Forest Implementation and Optimization for Big Data Analytics on LexisNexis’s High Performance Computing Cluster Platform,” Journal of Big Data, Springer, Vol. 6:68, 2019.

[8] V. Chinta, H. Kalva, and B. Furht, “Evaluation of Hadoop and HPCC for Multimedia Big Data Analysis,” Imaging and Multimedia Analytics in a Web and Mobile World Conference, 2017.

[9] 2022 Project Report for Award #2021585, NRT-HDR: Graduate Traineeship in Data Science Technologies and Applications, NSF, 2023.

[10] H. Kalva, V. Adzic, B. Furht, A. Krause, M.E.H. Elmon, and A. Perera, “Systems and Methods for Region Packing Based Compression,” Provisional patent 1097-072USP1, 2022.

[11] E. Cardenas, C. Shorten, T.M. Khoshgoftaar, and B. Furht, “A Comparison of House Price Classification with Structured and Unstructured Text Data,” Conference Proceedings of FLAIRS-35 Conference, 2022.

[12] H. Kalva, V. Adzic, B. Furht, A. Krause, M.E.H. Eimon, and A. Perera, “Systems and Methods for Video Packing, Encoding, and Decoding for Machine-Based Applications,” Provisional patent 1097-083USP1, 2023.

[13] H. Kalva, V. Adzic, J. Merlos, and B. Furht, “DCT Filtering for mAP in Video Coding for Machines,” Provisional patent 1097-086USP1, 2023.

[14] J. Merlos, H. Kalva, V. Adzic, and B. Furht, “Content Adaptive Multi-Scale Layer Filtering, Provisional patent 1097-0100USP1, 2023.

[15] C. Garbin, N. Marques, and O. Marques, “Machine learning for predicting opioid use disorder from healthcare data: a systematic review” Computer Methods and Programs in Biomedicine, https://doi.org/10.1016/j.cmpb.2023.107573, 2023.

[16] J. Marlos, H. Kalva, V. Adzic, and B. Furht, “DCT-based Filtering for Improving Machine Task Performance, Provisional patent, presented at the VCM Standard Committee Meeting, Geneve, Switzerland, July 2023.

[17] 2021585 Project Evaluation Report, NSF, 2023.