Winter January 31, 2019

2018 Florida Data Science for Social Good - Annual Report

Karthikeyan Umapathy
F. Dan Richard

Florida
Data Science
for Social Good

2018 FL-DSSG Annual Report

2/1/18

Table of Contents

Executive Summary

The Florida Data Science for Social Good (FL-DSSG) Program is a summer internship program that matches data science expertise with real-world problems in the public sector. The FL-DSSG program works with community organizations who are trying to affect change in their communities and who have data management, analytics, and visualization projects that have the potential to shift understanding around a community issue, influence planning, revise practices or see efforts in supporting community initiatives more focused or renewed. The mission of the FL-DSSG is to promote data-informed approaches and partners with organizations who use these approaches to solve wicked social problems while creating educational programs for aspiring data scientists.

During the 2018 program, the FL-DSSG received $23,000 funding from the University of North Florida. Funds supported stipends for 5 DSSG Interns, who worked 20 hours per week for 12 weeks and 3 DSSG interns, who worked 8 hours per week as well as registered for a Directed Independent Study course. Four non-profit organizations participated in the 2018 program: Baptist Health Y Healthy Living Centers (YHLCs), Family Support Services of North Florida (FSSNF), Girls Incorporated of Jacksonville (Girls Jax), and The Performers Academy (TPA). The Baptist Health YHLCs project objective was to identify the overall prevalence rate and the differences in clients who have metabolic syndrome versus those who do not. The dataset included biomarker, demographic, and personal health survey information. The prevalence of metabolic syndrome was found to be 26%, with these individuals reporting lower survey scores and poorer biomarker data than those who did not have metabolic syndrome. Surveys across time showed a marked increase in mean scores across several sections and no healthy individuals at a first screening developed metabolic syndrome by their next screening. Predictive modeling was used to find variables associated with getting metabolic syndrome to help wellness coaches in their evaluations. The goal of the FSSNF project was to help Family Support Services to develop a comprehensive data analysis process for their Purchase of Services (POS) system used for managing funding requests. We developed a process of data cleaning and preparation to help FSS better understand how to transform their data for analysis. We created Tableau dashboards and visualizations to give FSS the ability to investigate and gain insights from the data. The Girls Jax project data came from pre and post surveys that corresponded to six areas of programming in the middle school classrooms, and we were asked to identify if small incremental changes exist among the programming recipients. The analysis revealed that post-test scores were slightly higher for those schools that had the previous programming compared to those that did not, and girls are generally scoring better on post-test surveys and answering more favorable after programming. The goal of the TPA project was on measuring the impact of their art therapy programs. Their data primarily consisted of art pieces (paintings), poems, song lyrics, and exit surveys. The Latent Dirichlet Allocation (LDA) modeling was performed to identify key themes reflected in the poems and songs written by the program recipients. Due to the limited availability of objective data, FL-DSSG team proposed a new framework for collecting pre-test and post-test data.

On August 22nd, 2018, the FL-DSSG team presented the findings of these projects to the Jacksonville community in a culminating event called "Big Reveal." There were over 160 community members in attendance. The videos of the presentation can be found on the FL-DSSG YouTube channel (included at the end of this report). AgileThought, a software consulting company sponsored the 2018 Big Reveal event.

FL-DSSG partnership with the Nonprofit Center for Northeast Florida is integral to the success of the program. The internship program activities would not have been possible without the generous support, confidence, and guidance of the CEO of the Nonprofit Center, Rena Coughlin. We want to thank the staff of the Nonprofit Center for supporting the program efforts. Critical support for the program was provided by industry Sherpas (mentors), Faculty Leads, the FL-DSSG Advisory Board members, and UNF administrative and support staff. We also would like to thank student volunteers and Chirag Rana for taking pictures at the Big Reveal event. With these collaborative and cooperative efforts, we have and will continue to make a meaningful impact in our communities.

"We are social trustees of knowledge with a unique capacity to do social good."

About FL-DSSG

The Florida Data Science for Social Good (FL-DSSG) program blends data science and technology design to inform and solve important social problems in Northeast Florida. We accomplish this goal by partnering with local non-profit and governmental agencies, analyze their data, and find impactful solutions to social issues. The FL-DSSG program provides internship opportunities for students from diverse disciplines to work on these data science projects along with our community partners. FL-DSSG program offers local non-profit organizations data analysis and data visualization reports to support their work in the Northeast Florida community.

FL-DSSG is hosted at the University of North Florida (UNF). The FL-DSSG program is inspired and modeled after other successful DSSG programs in Chicago and Atlanta. The FL-DSSG program at UNF is being spearheaded by Dr. Dan Richard, Associate Professor in Psychology, and Dr. Karthikeyan Umapathy, Associate Professor in Computing. Over the past six years, Dr. Richard has been helping various non-profit/government agencies with data analysis as well as visualization and interpretation of the results. Dr. Umapathy has worked with local non-profit/government agencies for the past nine years in developing technology solutions to improve program efficiencies.

FL-DSSG program has formed a 2-member industry advisory board to seek guidance from experts with in-depth knowledge of data science and STEM workforce readiness. Advisory board members provide recommendations to program directors on project selection and recruitment of mentors,

help identify resources for the FL-DSSG program, assist with other relevant program activities, and advocate for the success of the FL-DSSG program. FL-DSSG advisory board membership is a voluntary service. 2018 FL-DSSG advisory board members are Robert March, Chief Technology Officer from NLP Logix, and Arri Landsman-Roos, Director of Analytics from Jacksonville Jaguars.

2018 FL-DSSG program was made possible by the generous support of $20,000 from the University of North Florida. Funds supported DSSG internship opportunities for students and other expenses associated with programmatic activities. The fund provided by the UNF helped in supporting eight DSSG interns impacting four nonprofit organizations.

Summer Internship

The FL-DSSG program is an intensive, twelve-week paid summer internship experience. DSSG interns are placed on multi-disciplinary teams and matched with mentors to address real-world problems for our clients in the Northeast Florida region. The 2018 FL-DSSG summer internship program started on the first week of June and ended on the third week of August. Interns received valuable experience with data management, analysis, technology, and community needs. DSSG program directors supervised interns, while interns received guidance from industry mentors as well as faculty project leads.

2018 DSSG Interns

For the 2018 program, we recruited six UNF students and two non-UNF students as DSSG interns, see listed below.

- Yekaterina (Kat) Bardash, Master of Science - Data Science, Regis University, Denver, CO
- Gabrielle Coker, Bachelor of Science - Mathematics, University of North Florida
- Crista Cummings, Bachelor of Arts - Anthropology, University of North Florida
- William Giroux, Master of Public Administration - Public Administration, University of North Florida
- Tabitha Powell, Master of Science - Psychological Science, University of North Florida
- Danish Sayed, Bachelor of Science - Statistics with minor in Chemistry, University of North Florida
- Habeeba Siddiqui, Master of Science - Computer Science, University of Illinois, Springfield, IL
- Britni Surprenant, Master of Science - Computer Science, University of North Florida

DSSG Intern Responsibilities

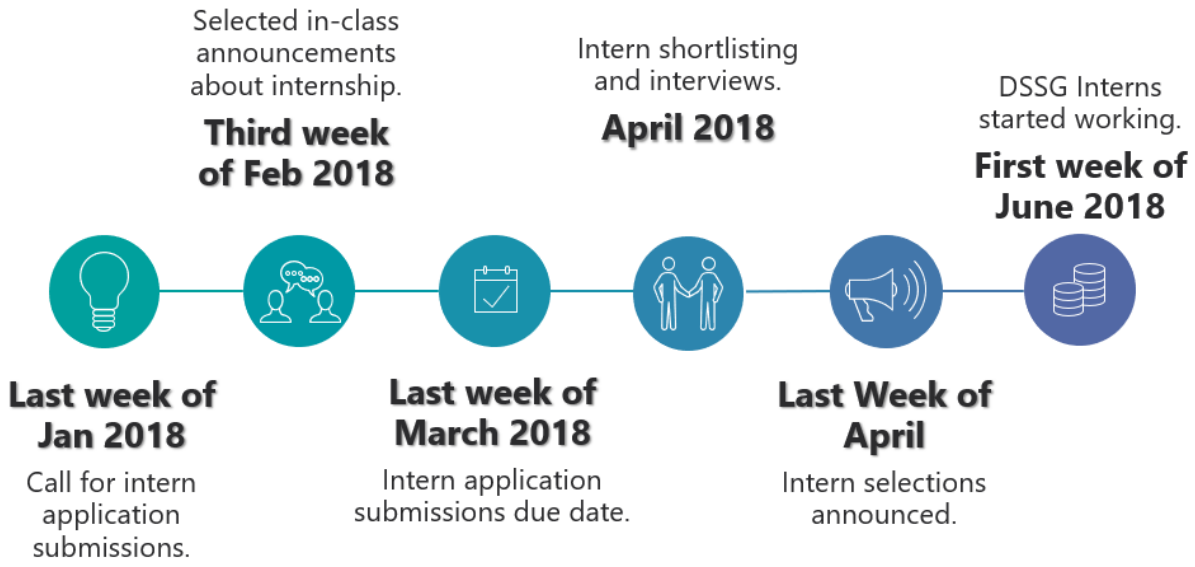
FL-DSSG interns were expected to provide a wide variety of data science services, including data mining, data analysis, data visualization, and assisting in data-driven decision-making. The work of the FL-DSSG interns include the following:

- Work towards understanding Community Partner needs and developing a data science product or report that addresses the identified problem;
- Help Community Partner understand the data-driven decision-making process used in relation to addressing the identified problem;
- Analyze the problem context and meaningfully document the identified problem;
- Document data specifications, analysis approaches, and findings associated with the project;
- Maintain regular communication with the Community Partner, Program Directors, Faculty Leads, and Professional Mentors (aka Sherpas) to obtain feedback on progress made;
- Collaborate with team members oriented toward the success of the projects using project management toolsets provided within the FL-DSSG program;
- Work within regularly scheduled timeframes of Monday, Tuesday, and Wednesday from 9:00 a.m. through 1:00 p.m. each day, and to work within flexible hours for the remaining 8 hours, within each week of the program;
- Notify Program Directors of any conflicts of interest, project impediments, and unethical behaviors within the scope of the FL-DSSG program immediately upon the occurrence of the relevant incident;
- Maintain confidentiality of data and products associated with the FL-DSSG program and maintain professional and ethical conduct with all parties involved; and
- Provide a concluding presentation that outlines all relevant findings and provides concluding recommendations to the Community Partner.

Selection Process

DSSG intern selection process started in late January 2018. We contacted instructors of courses that dealt with data mining, statistical analysis, and research methods in UNF, and requested instructors to post information about DSSG internship. Interested students submitted internship applications via Qualtrics questionnaire. By the end of March 2018, we received 22 applications. Based on the project needs and required skill sets for those projects, we interviewed 13 applicants during the first and second weeks of April. Based on the interview performance and information gathered, we offered an internship to 10 applicants, of which 8 accepted the offer. The 20-hours-per-week internship was offered to 6 applicants, and 8-hours-per-week internships with directed independent course option were offered to 4 other applicants. DSSG interns started their work on the projects during the first week of June 2018. See Figure 1 for the timeline followed for intern selection.

Figure 1: DSSG Intern Selection Timeline



Sherpas and Faculty Leads

Data Science is an evolving field; thus, the success of data science projects depends upon using current industry best practices. FL-DSSG relies on its industry Sherpas and faculty leads to provide industry best practices and guide DSSG interns to ensure the produced output is of quality and successfully creates intended effects for the client organization. Sherpa and faculty leads work along with program directors to mentor interns and provide technical guidance on the project tasks. In 2018 program, DSSG interns were mentored by five industry Sherpas and 8 faculty leads listed below. Sherpas and faculty leads meet DSSG interns on every two weeks basis for one hour and 30 minutes during the 12-week internship period. During these meetings, intern report progress made on the project, discuss challenges they face, and mentors provide their guidance on performing analysis, insights on what to look for, and shared tips on improving the project work.

2018 Sherpas (industry mentors) were:

- Naveen Agarwal, Principal and Founder at Creative Analytics Solutions, LLC
- Kellen Blumberg, Manager of Advanced Analytics at Jacksonville Jaguars
- Candace Dorn, Operations Analyst at Jacksonville Energy Authority (JEA)
- Jay Lewis, Digital Insights & Analytics Manager at TIAA Bank
- James Parks, Data Scientist at AgileThought

2018 Faculty Leads were:

- Dr. Emma Apatu, Assistant Professor, Department of Public Health, Brooks College of Health
- Dr. Beyza Aslan, Associate Professor, Director of Math Center, Department of Mathematics & Statistics, College of Arts and Sciences
- Dr. Georgette Dumont, Associate Professor, Department of Political Science & Public Administration, College of Arts and Sciences
- Dr. Lakshmi Goel, Professor, Coggin Endowed Strategic Professor, Department of Management, Coggin College of Business
- Dr. Julie Merten, Assistant Professor, Public Health Internship Director, Department of Public Health, Brooks College of Health
- Dr. Amanda Pascale, Assistant Professor of Higher Education Administration, Department of Leadership, School Counseling & Sports Management, College of Education & Human Services
- Dr. Gordon Rakita, Professor of Anthropology and Director of Academic Technology, Department of Sociology, Anthropology & Social Work, College of Arts and Sciences
- Dr. Sandeep Reddivari, Assistant Professor, School of Computing, College of Computing, Engineering, and Construction Management

Projects

The 2018 FL-DSSG program sponsored four community partner projects: Addressing Metabolic Syndrome (Baptist Health Y Healthy Living Centers); Patterns and Trends in Child Welfare Resource Systems (Family Support Services of North Florida); Breaking the Cycle of Poverty (Girls Incorporated of Jacksonville); and Empowering At-risk Youths through the Arts (The Performers Academy).

Project Selection Process

Partner projects are selected based on their fit with the data science approach and based on the potential impact of the project on the health and welfare of the community. The features of a successful partner project are:

1. **A wicked problem** – data science projects attempt to address a wicked problem - a vexing, persistent social or cultural issue that is complex in nature, interconnected with other problems, and requires many people working together to affect change. Projects could address wicked problems in one of several areas, including education, the environment, government services, health care, healthy living, safety/security, smart urban development, social & economic inequality, and sustainability.
2. **A committed team** – attractive project partners will have organizational leaders who are committed to solving these wicked, real-world problems and who are supportive of using evidence to make important decisions around social good initiatives. Project partners will provide a dedicated staff person in the organization who is willing to work weekly with data science students and mentors on the data science project.

3. **Relevant data and lots of it** – An essential component of any data science project is, obviously, data. Data can take many forms. The project data must represent the complexity of the wicked problem. Data might come from the project partner organization or from other publically available sources, or both. The size and complexity of the data must provide an adequate experience for students to learn data science techniques and to have relevance for addressing the wicked problem.
4. **A landing pad** – DSSG projects generate products that are useful for participating organizations, partners, and communities. These products often go beyond a single report or project summary. The FL-DSSG Directors and Mentors will work with participating project partners to ensure that the organization has the computing, data, and staff infrastructure to support the delivery of the data science project products.

In collaboration with the Nonprofit Center of Northeast Florida, the FL-DSSG program hosted a partner information session to provide potential program partners with information about the program and on how they could apply for project selection. Potential partners completed an online application, and the FL-DSSG Directors conducted interviews with all applicant organizations. The final list of four organizations was selected based on the suitability of the project and partner based on the timeline, resources, and goals. All project partners were selected by May 2, 2018. Figure 2 provides a timeline of the partner selection process.

Figure 2: Partner Selection Timeline

Baptist Health YHLCs Project

The Baptist Health and YMCA partnered to create Healthy Living Centers (YHLC) with a focus on community outreach and preventive care. Their wicked problem centered around the country-wide issue of metabolic syndrome. Specifically, the overall prevalence rate and the differences in clients who have metabolic syndrome versus those who do not. This project focused on data from Applied Health Analytics which included biomarker, demographic, and personal health survey information. The prevalence of metabolic syndrome was found to be 26%, with these individuals reporting lower survey scores and poorer biomarker data than those who did not have metabolic syndrome. Surveys across time showed a marked increase in mean scores across several sections and no healthy individuals at a first screening developed metabolic syndrome by their next screening. Predictive modeling was used to find variables associated with getting metabolic syndrome to help wellness coaches in their evaluations. Overall, the YHLCs hope to use this information going forward to

provide better programming, individual care, and help reduce the prevalence of metabolic syndrome across Duval County.

Family Support Services Project

Family Support Services of North Florida (FSS) is the leading agency for foster care, adoption, and child welfare in Nassau and Duval counties. The goal of the 2018 Florida Data Science for the Social Good program was to help FSS develop a comprehensive data analysis process for their Purchase of Services (POS) system used for managing funding requests. To accomplish this goal, we documented our process of data cleaning and preparation to help FSS better understand how to transform their data for analysis. Additionally, we created Tableau dashboards and visualizations to give FSS the ability to investigate the data they shared with us and adapt them for use with future data. Our recommendations revolve around adapting the POS system to meet FSSs needs better.

Girls Inc. Jacksonville Project

The wicked problem presented by Girls Inc. of Jacksonville was to end cycles of female generational poverty. Girls Inc. of Jacksonville tackles this by making small incremental changes in the elementary and middle schools through Smart, Bold, and Strong programming. The data provided to Florida Data Science for Social Good (FL-DSSG) was in the form of pre and post surveys that corresponded to six areas of programming in the middle school classrooms. Although we were not truly able to answer the wicked problem with the data provided, we focused on analyzing survey data to identify if small incremental changes exist. One basic finding was that post test scores were slightly higher for those schools that had previous programming compared to those that did not. Another basic finding is that girls are generally scoring better on post-test surveys and answering more favorable after programming. Recommended next steps primarily focus on the data collection process as well as suggestions to better address the wicked problem of ending cycles of female generational poverty.

Performers Academy Project

The Performers Academy is an art therapy institution that aids troubled youth. The Performers Academy had the goal of measuring the impact of their art therapy programs. Their data primarily consisted of art pieces (paintings), poems, song lyrics, and exit surveys. The goal of the project was to help The Performers Academy demonstrate the impact of their programming on troubled youth. Toward this goal, the data science question was to translate the focus of the program, to see troubled youth develop healthy forms of expression of their trauma through art and to develop trust in others for healthy relationships, into actionable and accountable measures of impact. The analytical approach included finding a "theme" in the poetic and song lyric data through Latent Dirichlet Allocation (LDA) modeling. The team also proposed a new framework for collecting pre-test and post-test data.

The Big Reveal Event

FL-DSSG hosted 2018 Big Reveal event on Wednesday, August 22 at the WJCT Studios in Jacksonville. The event marked the culmination of the summer internship program offered by UNF in collaboration with the Nonprofit Center. At the event, FL-DSSG interns presented findings and reveal insights gained from the 2018 DSSG projects. The event was attended by over 160 individuals from nonprofit sector. The 2018 Big Reveal Event was sponsored by AgileThought, a software consulting firm. Project presentation was video recorded and it can be viewed at the FL-DSSG You Tube channel at: https://www.youtube.com/watch?v=V0naHOlwIL0.

Conclusions

The ability to affect change and do good in one's community increasingly depends on having the right information at the right time to make the right decisions about things that are most important. Directors of community programs as well as funding agencies want evidence of impact and demonstrated efficiency in programs that serve our communities. Often, the information available to meet these needs are not well organized, not well understood, and not packaged in a way that helps those working in the community do their best.

The FL-DSSG program was formed to address a lack of data scientist capacity within the nonprofit sector and to use these data science projects as an opportunity to train students from diverse backgrounds to be a data scientist. In the year 2018, the FL-DSSG program worked with four non-profit agencies in the North Florida region to support their data science needs in service to addressing the region's persistent social problems. In addition, eight interns were trained to use data scientist skillsets for solving social good problems.

There is a growing demand for trained data scientists who can develop data-driven problem solving approaches to resolve our social issues. Through the offering of FL-DSSG program, we learned the extensive need of this program for our region but also there is vested interest in the community for this program to succeed beyond. We were fulfilled in discovering the mutual interest and benefit our community partners, industry partners, faculty, and students experienced through the program. Together we worked as social trustees of knowledge, each of us contributing our wisdom and time to address pressing social issues in our community.

Get More Information

For additional information about the FL-DSSG program; its process, methodologies, tools, and techniques; and how to participate with the program contact FL-DSSG program directors. Visit FL-DSSG website for contact information.

2018 FL-DSSG program was funded by the Nonprofit for Northeast Florida and the University of North Florida.

2018 Big Reveal Presentation Resources

Project Presentation YouTube Video: https://www.youtube.com/watch?v=o4VhuP8vDn0

2018 Big Reveal Presentation Slides: http://bit.ly/18FLDSSGReveal

2018 Big Reveal Event Pictures:

https://www.facebook.com/pg/FLDSSG/photos/?tab=album&album_id=515066478955218

2018 FL-DSSG Interns

Yekaterina (Kat) Bardash, Master of Science - Data Science, Regis University, Denver, CO
Gabrielle Coker, Bachelor of Science - Mathematics, University of North Florida
Crista Cummings, Bachelor of Arts - Anthropology, University of North Florida
William Giroux, Master of Public Administration, University of North Florida
Tabitha Powell, Master of Science - Psychological Science, University of North Florida
Danish Sayed, Bachelor of Science - Statistics with minor in Chemistry, University of North Florida
Habeeba Siddiqui, Master of Science - Computer Science, University of Illinois, Springfield, IL
Britni Surprenant, Master of Science - Computer Science, University of North Florida

FL-DSSG Program Directors

Dr. Dan Richard, Associate Professor, Psychology Department

Dr. Karthikeyan Umapathy, Associate Professor, School of Computing

Florida Data Science for Social Good (FL-DSSG)

University of North Florida

December 2018

www.dsgg.unf.edu

www.facebook.com/FLDSSG/