

### FALL 2019

# DEPAUL CENTER FOR DATA SCIENCE

### IN THIS ISSUE

Facts & Stats Alumni Information Students & Faculty News Meet Our New Faculty



DEPAUL UNIVERSITY COLLEGE OF COMPUTING AND DIGITAL MEDIA

## FACTS & STATS



- The Center for Data Science (formerly known as Center for Data Mining and Predictive Analytics) was launched in Spring 2010 under the leadership of three College of Computing and Digital Media faculty and one Kellstadt Graduate School of Business faculty
- The MS in Data Science program was started in Fall 2010 under the name of MS in Predictive Analytics
- Top industries hiring CDM Data Science students include IT/Technology, Insurance, Education, and Food/Beverage
- Top job positions held by students after graduation include Data Scientist, Data Analyst, Data Engineer, and Consultant
- Students think Data Analysis and Regression, and Fundamentals of Data Science are among the most useful courses in the program; other helpful courses include DSC 450 and DSC 540

### **ALUMNI INFORMATION**

### **MS IN DATA SCIENCE OUTCOMES**

2010 - 2018 400 300 Students Enrollment 200 100 Alumni Network 0 2010 2011 2012 2013 2014 2015 2016 2017 2018

#### **TOP INDUSTRIES**

- IT/Technology
- Insurance
- Education
- Food/Beverage
- Consulting
- Insurance
- Government

#### **REPRESENTATIVE POSITIONS**

- Data Scientist
- Data Analyst
- Data Engineer
- Consultant

### **STUDENTS & FACULTY NEWS**

- In July, the School of Computing hosted a Data Science BootCamp for Chicago Public School high school students. The initiative was in partnership with the CPS Office of Computer Science, DePaul University and University of Massachusetts - Boston.
- Jillian Stoller-MS Data Science- was among the 200 top college students from across the country selected to participate in ThinkChicago, a behind-the-scenes tech event coinciding with Lollapalooza.
- MS Data Science student Sierra Sellman and her teammate Michelle Rodrigue, a graduate student at the Georgia Institute of Technology, won first place in the US Department of Energy's 2019 Geothermal Design Challenge.
- MSDS students and faculty have new publication Yan Yu, Rob Weddell, Jacob Furst, Daniela Raicu. "Utilization of Instance Hardness in Multiple Annotations Active Learning". ACM SIGKDD Data Collection, Curation, and Labeling for Mining and Learning Workshop. 2019
- MSDS students and faculty have new publication Yan Yu, Yiyang Wang, Jacob Furst, Daniela Raicu. "Identifying Diagnostically Complex Cases through Ensemble Learning". *International Conference on Image Analysis and Recognition. 2019*
- Dr. Tanu Malik was awarded the NSF Career Award "Advanced Containers for Reproducibility in Computational and Data Science", June 1st, 2019 – May 31st, 2024 (Amount: \$498,888.00).
- Dr. Daniela Raicu and Dr. Jacob Furst were awarded a NSF REU supplement to host the Medical Informatics Experiences in Undergraduate Research Principal Investigators Meeting, August 2nd, 2019 - February 29th, 2020 (Amount: \$74,396.00).

# **NEW FACULTY**



Dr. Roselyne Tchoua is an Assistant Professor at the School of Computing, DePaul University. She received her PhD in computer science from the University of Chicago, focusing on Hybrid Human Machine Scientific Information Extraction. Her interests gravitate around making seemingly inaccessible technology or manageable amounts of data more reachable.

#### What are your current research interests?

My current interests revolved around scientific information extraction. I enjoy working with other disciplines to extract scientific insight from data using crowdsourcing, machine learning and natural language processing.

#### Why did you choose this profession/field?

I chose this field because I really enjoyed my position in the scientific data group at Oak Ridge National Laboratory. I wanted to continue working as a computer scientists in interdisciplinary research and development.

What courses will you teach and what benefits do you see in these courses?

I will be teaching data science courses starting with machine learning applications in the Autumn quarter. I think this is a really exciting field in computer science, there is so much information all around us, but none matters unless we can find ways to make sense of it. According to the Economist, the most valuable resource is now data, therefore skills related to manage, visualize and interpret data are becoming more and more important.

#### What is your teaching style like?

I am a firm believer that it is as critical for students to acquire broadly applicable skills than mere knowledge of the material itself. It is important that students develop an intuition for problems based on a strong theoretical foundation. I strongly encourage critical thinking and understanding by using examples, hands-on exercises, in-class questions and discussions.

#### What are your research plans for this year?

I plan to work on a couple results and ideas related to my PhD project i.e. extracting scientific named entities and relations from the literature. I also would like to explore new applications of my research areas working with other CDM faculty.

## **NEW FACULTY**



Thiru Ramaraj is an Assistant Professor at the School of Computing, DePaul University, Chicago, USA. He received his BE in Computer Science & Engg. from Annamalai University, India, MS & PhD in Computer Science from Ball State University and Montana State University respectively. His research focus is mainly on Computational Biology and applied bioinformatics.

#### What are your current research interests?

My research interests are focused on the development of novel computational approaches and infrastructures required to analyze, manage, mine and integrate big biological data. Currently I have three active NSF funded projects, (i) Investigating graph data structures to effectively represent pan-genomes and associated algorithms for analyzing pan-genomes, (ii) Biofilm Resource & Information Database (BRaID): A Tool to Fuse Diverse Biofilm Data Types, and (iii) Re-sequencing several accessions of cotton to address fundamental biological questions of polyploidy genomes and the genetic diversity of cotton.

#### Why did you choose this profession/field?

With my training and background in Computer Science, I decided to pursue research in the field of Computational Biology/Bioinformatics because I was fascinated by the interdisciplinary nature of the subject that requires integrating biology and computer science.

#### What courses will you teach and what benefits do you see in these courses?

I will be teaching classes that are primarily part of the MS in Data Science program. DSC424 or DSC441 are among the classes that I will be teaching in my first year. In a nutshell these classes will attempt to equip students with the underlying theory and methods/techniques/skills needed to manage, mine, integrate, create knowledge, visualize whatever big data they will encounter in the real world.

#### What is your teaching style like?

My teaching philosophy would be to engage students and guide them to cultivate knowledge, problem-solving skills, and inculcate curiosity and passion for learning. I truly believe that the best way to effectively keep students engaged in a class room setting is a combination of both lecture-based and hands on practical learning strategies. Teachers have a great responsibility of not only making sure that students understand the subject material in detail but also ensure they are able to apply what they learn in real world applications.

#### What are your research plans for this year?

I hope to have my current NSF grants transferred to DePaul by the end of the Autumn quarter and have students involved in the different research tasks associated with these projects. I am currently working on a proposal titled "Pangenome-wide association studies" with my collaborators at the NCGR and my alma mater, Montana State University - Bozeman. I would like to submit a NSF CAREER award by the end of my fist year, very broadly, focusing on pioneering novel computational approaches for problems related to heterogeneous, voluminous biological data. Lastly I am excited for potential research collaborations with my colleagues at the School of Computing here at DePaul University.

#### **DePaul Center for Data Science**

## **NEW FACULTY**



Ilyas Ustun is an Professional Lecturer at the School of Computing, DePaul University, Chicago, USA. He earned his PhD from the Department of Modeling, Simulation, and Visualization Engineering at Old Dominion University in Norfolk, VA. His research interests include Data Science, Big Data Analytics, Data Visualization, Data Mining, Machine Learning, and Deep Learning.

#### What are your current research interests?

My current research interest involves machine learning and data analysis in the broader fields of transportation and metabolomics. That may sound weird, as these fields are far too separate from each other. But the common thing is "data", and at the end it is all about analyzing data and making it useful. I am involved in several projects like work zone mobility and safety analysis, traffic corridor mobility analysis, using LIDAR data for freight analysis and planning. On the metabolomics side, I am involved in projects regarding the diagnosis of a person with Alzheimer's or Parkinson diseases. The predictive models aim to correctly identify the presence of these illnesses in patients by making use of data collected via controlled experiments.

#### Why did you choose this profession/field?

I love anything related to data. What something started as in interest evolved into passion over time. After years of searching for the one thing I like, and trying many different things, I believe I finally have found my area of interest which is data science.

#### What courses will you teach and what benefits do you see in these courses?

My first courses at DePaul will be DSC 425 Time Series Analysis and Forecasting and DSC 341 Foundations of Data Science. Both courses are very important in data science. Time series analysis forms the backbone of financial data. Any data which is collected over time like traffic speed, weather and so on can be analyzed using time series models. Being able to learn from data and forecast what will happen next has a lot of potential. Foundations of data science will be an introduction into the field of data science. We will build the very foundations of data analytics. Having a good foundation is essential to be successful later in more advanced courses.

#### What is your teaching style like?

My teaching style involves active engagement with students and keeping their interest alive. The best way to keep the interest up is to connect the class material with real life.

#### What are your research plans for this year?

My research plans are to continue with what I am doing now and to collaborate with the valuable faculty at DePaul University. DePaul Center for Data Science provides great opportunities for me to be involved in the various projects. Together with the faculty and the students I believe there are many nice projects waiting for me at DePaul.

## EDUCATIONAL

To nurture the growth of the next generation of data scientists and computer scientists to better prepare them for data-related computing careers. Students work on state-ofthe-art research and practice activities under the supervision of faculty members.

### CONTACT US

dsc@depaul.edu http://cds.cdm.depaul.edu/

https://twitter.com/DePaulCDS

243 S Wabash Ave, Chicag<mark>o</mark>, IL 606<mark>04</mark>