

WINTER 2020 DEPAUL CENTER FOR DATA DATA SCIENCE

IN THIS ISSUE

New Year Greetings News Job Opportunities Student Events Meet Our Students New Projects



DEPAUL UNIVERSITY COLLEGE OF COMPUTING AND DIGITAL MEDIA

NEW YEAR GREETINGS FROM THE CENTER OF DATA SCIENCE DIRECTOR

Dear Students,

Happy new year!

2020 marks the tenth anniversary of the DePaul Center for Data Science. The Center was created in Spring 2010 and launched together with the MS in Data Science (DS) Program (formerly known as Predictive Analytics). Over the past ten years, the



Center continued to encourage and support students to engage in capstone projects, independent studies, and research credits – all aiming to help students build a strong foundation for data-oriented careers in computing.

We powered through another exciting year full of achievements. Our faculty Dr. Tanu Malik and Dr. Jacob Furst received grants for their hard work in computer systems and medical fields, respectively. We also welcomed new professors to the School of Computing, in particular to the Data Science Program. Dr. Ilyas Ustun, Dr. Roselyne Tchoua, and Dr. Thiru Ramaraj joined our Data Science faculty team and brought new projects and research opportunities to our students. Our Data Science student group organized 15 events to help students build networking skills, practice interview and enhance coding abilities. Our alumni network also provided over 30 data science job openings across numerous industries.

Going forward into the new year, we will stay true to our mission to nurture the growth of next-generation data scientists and computer scientists through state-of-the-art curricula, interdisciplinary research projects, and industry collaborations. Please feel free to reach out to our faculty to get engaged in any of the activities organized by the Center, student organization, and projects.

I am confident that 2020 will be a year full of rewarding experiences. Thank you for being a member of DePaul Data Science.

Daniela Raicu, PhD Director of the DePaul Center for Data Science

NEWS

Conference News

- Robotic Process Automation, Artificial Intelligence & Cognitive Conference -Chicago will be hosted by Techies Meetup on March 6, 2020
- Transforming Data With Intelligence(TDWI) World Conference will be hosted by TDWI from May 10 - 15, 2020 at Hilton Chicago
- Data Science Conference by NeoNorm will be taking place from May 14 15, 2020 at the University of Chicago Gleacher Center

Students & Faculty News

- On September 10, 2019, Jason Chuah, Madeline Deeds, Tanu Malik, Youngdon Choi, and Jonathan Goodall attended the PARCO Symposium on Tools and Infrastructure for Reproducibility in Data-intensive Applications in Prague, CZ and presented on their topic, Documenting Computing Environments for Reproducible Experiments
- Since June 2019, DePaul faculty Dr. Sheena Erete and students in CDM have been working with the Chicago Mayor's office and Northwestern University on Every Youth Connected, a project that aims to leverage out-of-school learning opportunities and city resources to achieve more equitable youth participation. By creating new models for data-focused and equity-centered community collaboration, this work has great potential to impact the local and national levels
- Dr. Jacob Furst and his collaborator Dr. Leonard Jason in the College of Science and Health (CSH) received a National Institutes of Health (NIH) Grant of \$2,857,865 over a five year span from 01/01/2020 – 11/30/2024 on the topic Maintenance and Incidence of ME/CFS following Mono. Their project studies clinical and biological predictors of ME/CFS (also known as chronic fatigue syndrome) following intramuscular treatment to determine the incidence of ME/CFS in a sample of young adults

JOB OPPORTUNITIES





Our job opportunities come from:

Aetna	Deloitte	Gallagher	Northwestern Medicine
Apkarian Lab	Department of Police	IBM	P&G
AskWhai	Digital Capability Center	IRI	PCPI
CCC	Eight Eleven	KPMG	Quaxigma
Clearcover	First Trust	Northern Trust	SAP

STUDENT EVENTS

2019 Student Events hosted by Data Science Group

Networking

- Network with Professionals from Isobar (October 2019)
- Data Science Networking (September 2019)
- Google Tour (May 2019)
- Networking Event (February 2019)

Interview Skills

- Mock Interview Opportunity with Booksource (November 2019)
- Massive Technology Career Fair Preparation (October 2019)
- Professional Development (April 2019)
- How to Succeed in a Data Analytics Interview. Part 2 (April 2019)
- How to Succeed in a Data Analytics Interview. Part 1 (April 2019)

Coding Skills

- Chicago ML Workshop(collaboration with DDSG) (May 2019)
- Data Visualization (May 2019)
- Hands-on SQL (May 2019)
- Modeling with Python & GitHub (May 2019)
- Data Analytics Code Challenges. Part 2 (May 2019)
- Data Analytics Code Challenges. Part 1 (April 2019)

MEET OUR STUDENTS



Daniel Gutstein

I'm a native of Lincolnwood, Illinois, and did my undergraduate studies in Economics and Communications at Northeastern Illinois University. I began the M.S. in Data Science (back then it was called Predictive Analytics) program at DePaul in Fall 2016. I've been a full-time student here since, and I've taken the course-load of machine learning and computer vision classes. When I'm not in class or performing research with the Medical Informatics and WHEEL laboratories on the 7th floor of the CDM building, I spend my time engaged in Talmudic study and entertaining my nephews and nieces.

Can you briefly introduce your research?

My work for the Medical Informatics and WHEEL laboratories involves collaborating with my peers and advisors to implement computer vision and audio extraction techniques that extract measurements from video recordings of clinical interactions. After extracting the features, we have used machine learning methodologies to make predictions regarding features of the interactions (such as who is speaking in a given frame). The long-term goal of the project is to create an automated, interactive feedback system for clinicians that can contribute to both improving patient care and reducing physician burnout.

Why did you choose a master's thesis as your capstone option?

My advisor - Dr. Daniela Raicu - was instrumental in encouraging me write a thesis as a way of summarizing all of the hard work we had done until that point. It is true that we had two papers which had already been published at international bioinformatic conferences, but doing the thesis opened up a new frontier for documenting and expanding the intricate, multi-layered processes of research and implementation which the project has entailed. Ultimately, of all the capstone options, I was excited by both the challenge and learning opportunity offered by writing a thesis, which is why I chose this route.

What have you learned from completing your thesis?

The process of researching, implementation, and writing is not for the faint of heart. Yet, I would say the most important thing I learned is the importance of group collaboration in order to facilitate the process of research from beginning to end. Although my name is on the thesis, I am very grateful to my thesis committee - Dr. Daniela Raicu, Dr. Enid Montague, and Dr. Jacob Furst - as well as to my parents, for helping me reach the finish line.

What advice do you have for students who are contemplating doing research?

Don't take any shortcuts, and don't be afraid to fail and redo your work. As Dr. Furst put it, in research, we do a lot of "re".

MEET OUR STUDENTS



Hao Wu

I am a first-year graduate student in data science with a concentration in computational methods. I hold a master's degree in mechanical engineering. Before attending DePaul, I have one year of engineer experience in the semiconductor area. I currently work as a graduate assistant at the DePaul Center for Data Science.

What can you say about your experiences in the MS data science program?

I really enjoy this program so far. As someone just stepping into the field of data science, the program provides a range of courses from fundamental to advanced data science concepts and techniques that have helped me quickly learn the basics while building up a foundation for more complex topics. Knowledgeable professors combine their industry experience with their teaching experience to equip students with necessary skills upon completion of the program.

Can you describe the DePaul Center for Data Science?

I discovered the Center when I was applying for the data science program, and it was one of the reasons I chose to attend DePaul. Therefore, I feel lucky to have been selected for the graduate assistant position at the Center. In this position, I am seeing how the Center operates behind the scene. In my opinion, the Center benefits students in three main ways. Workshops and networking events give students a chance to get in touch with companies seeking talent; job opportunities from our expanded network ensure a high employment rate; and research and capstone projects allow students to connect lectures and assignments with real-world problems.

NEW PROJECTS

Traffic Crashes (from Dr. Ilyas Ustun)

Description: Traffic crashes have a significant impact on the economy both in the form of property damage and also in the form of lost time. The congestion likely to happen in busy areas will cause waste of gas and air pollution. The worst are fatalities or severe injuries. The most vulnerable population in traffic crashes are pedestrians and cyclists. Identifying the crash-prone locations will help traffic safety, transportation planning, and law enforcement to prioritize their efforts and resources to minimize the risk of accidents. The Michigan traffic crash data is rich and contains a lot of information of the incident, severity, time, involvement of pedestrian or bike along other factors. There are 300K rows each year for the past 10 years. Analyzing this data will provide insight into many aspects of traffic crashes. Joining with other data sources can improve the quality of the data. There can be several paths that can be explored here, one of which is the how the pedestrians or bikes are affected when involved in a traffic crash. Other can be related to the propensity of drug usage increasing the severity of a traffic crash. There is a lot of data cleaning, wrangling, visualization, and modeling involved. By participating in this project the student will learn how to perform in-depth data analysis and apply different machine learning models to a real-life data set

Traveling Equipment (from Dr. Ilyas Ustun)

Description: This project investigates the problem of planning for the allocation of resources to provide services to spatially dispersed customers from a single or a network of hub locations, where resources are stored. These resources may be equipment such as seating and staging equipment in the entertainment industry, freight distribution vehicles, and trade-show booths. The resources may also be human resources such as consultants. One key decision for these problems is the dynamic reallocation of the resources to the hubs, where these resources are routed based on possible demand fluctuations. We consider this problem in the context of seating and staging services in the entertainment industry, where the planning process is guite complicated as it involves managing thousands of pieces of modular equipment among warehouses and events, such as concerts and sports games. We investigate improvement opportunities in terms of inventory and transportation management. This study is conducted in collaboration with SGA Production Services which provides temporary seating and staging for large-scale entertainment events with its network of six warehouses in the USA, where they store their equipment. The project has two aspects. One aspect is to analyze the data thoroughly and gain insights followed by predictive modeling and machine learning. The other aspect involves optimization of the equipment allocation which requires significant knowledge in operations research. By participating in this project the student will learn how to perform in-depth data analysis and apply different machine learning models to a real-life data set.

NEW PROJECTS

Human-in-the-loop Image Pattern Detection (from Dr. Roselyne Tchoua)

Description: New materials can provide solutions for key challenges in sustainability, e.g., in energy, new catalysts for more efficient fuel cell technology. One of the several challenges in new materials discovery is the identification of the crystalline phases of inorganic compounds based on an analysis of high-intensity X-ray patterns. Identifying these phases is equivalent to finding the crystal structures (arrangement of atoms) of new compounds, which then leads to determining their properties. Fully automated phase identification is challenging as the images generated with the X-ray instrument can be noisy and the patterns (or series of matching peaks) has to be identified across sets of multiple samples of materials. In this project, we explore a combination of visualization techniques, guided by humans to accelerate crystalline phase identification. We will build on previous work to develop automated pattern detection and investigate opportunities to integrate expert and non-expert feedback.

Hybrid Human-Machine Information Extraction

(from Dr. Roselyne Tchoua)

Description: Materials informatics is an emerging field that has the potential to dramatically reduce the time-to-market and development for new materials; computational models scan large datasets to identify candidates for new materials. As such methods rely on access to large, machine- readable databases, the traditional text-based physical handbooks will not suffice. However, there are few examples of these scientific digital databases and constructing new databases is a monumental and costly task requiring years of expert labor, as the data that populate these databases must often be extracted manually from free-text publications. While, machine learning efforts have begun in materials science, the lack of annotated text hinders attempts to leverage approaches developed for biomedicine for example. In this project, we will build on previous work which leverages human and automated approaches to extract scientific named entities from text. We will enhance this work to tackle scientific entity relation extraction. Specifically, we will explore comparable human-in-the-loop extraction approaches to continue to contribute to existing datasets of annotated materials entities and properties.

RESEARCH MISSION

To promote collaborative research among DePaul faculty in a variety of areas, including machine learning, text mining, NLP, health informatics, image analysis, recommender systems and database systems.

CONTACT US



dsc@depaul.edu http://cds.cdm.depaul.edu/ https://twitter.com/DePaulCDS 243 S Wabash Ave, Chicago, IL 60604

DePaul Center for Data Science