Data Intensive Studies Center (DISC) at Tufts

Abani Patra
Director DISC and Stern Family Professor of Mathematics and Computer Science,
Ballou Hall, Tufts University, Medford MA 02155
Abani.patra@tufts.edu

I. Tufts University

Tufts is a student-centered research university dedicated to the creation and application of knowledge. We are committed to providing transformative experiences for students and faculty in an inclusive and collaborative environment where creative scholars generate bold ideas, innovate in the face of complex challenges and distinguish themselves as active citizens of the world. Tufts is a comprehensive university in the framework of a liberal arts university.

II. Data Intensive Studies Center (DISC)

Data Intensive Studies Center is Tufts' newly instituted center to generate and foster data intensive research, education and scholarship at Tufts University. DISC strives to increase volume of data science research and to provide high quality expertise and consultation to the faculty and researchers at the University. In addition, DISC aims to nurture and enhance a culture of data literacy and awareness by developing initiatives to create and strengthen networks, increase faculty and student engagement across the University. DISC seeks to cultivate the University's flourishing intellectual community and

> **Goal:** *The data intensive Studies Center (DISC) will lead in **creating, curating and educating** in data science knowledge, skills and reusable intellectual property in all areas of study at Tufts.*

research activities that utilize emerging transformative cross-cutting data science techniques like machine learning and Artificial Intelligence. We recognize the critical role of DISC to fill key gaps and create unique ecosystem to increase data science research at Tufts University. DISC is integrated with the Provost's office and actively promoted and supported by the Office of the Vice Provost for Research(OVPR).

III. Why DISC Now?

Data driven modeling and the impact of reliable model based prediction to support sound policy making has been reinforced every day during this pandemic. It is also increasingly clear that while the fundamental modes of decision making whether in scientific inquiry or important social/economic policy making are now much more data centric, our ability to deal with large, complex, incomplete and noisy data for use in reliable prediction is an open grand challenge. The rapidly growing methodologies of *AI and machine learning have much promise but their use in applications of high consequence* is still in its infancy. Tufts investments in the DISC are essential and timely to ensure Tufts community is at the forefront of this revolution. Our investments now will yield high returns. DISC will focus each year on particular themes in the broad area of data sciences.

IV. Research Thrusts and Major Grant Development

The creation of DISC has catalyzed several new research and education thrusts at Tufts. Major new grants supported by Tufts already in award include an NSF TRIPODS (PI L. Cowen), DOE PSAAP Center, NSF CSSI, NSF CC*, NSF/NRT. To support the development of major new data science we have developed two mechanisms for research initiation – *seed grants* and *study groups*. The seed funds will support preliminary work for new data science and data science +X research. The study groups include domain and data scientists to explore a "broader area" with support for group activity, speakers and a "sandbox".

Outcomes should be potentially transformative new approaches. DISC faculty and scientists time will be allocated to both programs.

V. Education and Training*:* In partnership with several units (TTS and CTSI notably) we started the Tufts Initiative for Data Analytics and Learning (TIDAL). TIDAL courses launched at the beginning of the COVID-19 related campus shutdown and have been successful with participation averaging 23 students per course for over 35 classes ranging in length from 1hr to 15hrs. The initial offerings were skills oriented and targeted idled staff and faculty with virtual data science courses but have gone on to much popularity among students at different levels and research staff. Topics have ranged from a Python bootcamp to Matrix Methods for Machine Learning. At Tufts several MS programs have also started to support the need for data science education. However, no PhD program exists to create the thought leaders in this emerging discipline. DISC has started developing such a cross-disciplinary program.

VI. Community*:* DISC will seek to form strong partnerships with local communities. This year we have engaged in a COVID-19 tracking project for the town of Medford providing data science resources.