BUET ACM CHAPTER

presents

Data Science for "All"

Speaker: **Shubhra Kanti Karmaker (Santu)**

Assistant Professor, Department of Computer Science and Software Enginering,
Auburn University, USA

Shubhra Kanti Karmaker ("Santu") is a Tenure-Track Assistant Professor in the department of Computer Science and Software Engineering at Auburn University, Alabama. His research interest lies at the intersection of Big Data, Artificial Intelligence and Natural Language Processing. Before joining Auburn University, Santu was a Postdoctoral Research Associate in the Laboratory for Information and Decision Systems at Massachusetts Institute of Technology (MIT), hosted by Dr. Kalyan Veeramachaneni. Prior to joining MIT, he completed my Ph.D. in computer science from University of Illinois Urbana Champaign (UIUC) under the supervision of Prof. ChengXiang Zhai. During his PhD, he also worked as a research intern at Microsoft Research (Summer 2017 and 2018), Yahoo Research (Summer 2016) and @WalmartLabs (Summer 2015). Santu has published numerous research papers at premier conferences including ACM SIGIR, WWW, ACM CIKM, IEEE CEC and ACL SIGNLL CoNLL. Santu has served as a program committee member in SIGIR 2020, ACL 2019, WSDM 2020, WSDM Demo Track [2019, 2018], IUI 2019 and also served as invited reviewer of IEEE Transactions on Knowledge and Data Engineering (TKDE) and Neurocomputing journals. Before joining the Ph.D. program, Santu served as a full-time lecturer of Computer Science and Engineering at BUET during 2012-2014.

Abstract

Big data is ubiquitous across domains, and more and more stakeholders are choosing to use machine learning to get the most out of their data. This pressing need has inspired researchers to develop tools for Automated machine learning (AutoML), which is essentially automating the process of applying machine learning to real-world problems. But although automation and efficiency are some of AutoML's main selling points, the process still requires a surprising level of human involvement from a Data Scientist and are still far from a "real automatic system". As a result, AutoML tools are not yet directly usable by domain experts like doctors, business professionals, social scientists and so on, who have little / no knowledge of machine learning / data science. In summary, Data science is not yet open to "all". How can we make data science more accessible to the general people? This talk will focus on this big question while discussing three independent yet related line of works. The first direction is about helping general users annotate big corpus of text data with minimal guidance from the users. The second direction is about enabling end users perform interesting semantic analysis of unstructured data without worrying about the underlying intricate details of machine learning. The third and final direction will be laying out a vision for Virtual Interactive Data Scientist, a natural dialog based intelligent agent which can be thought of as the future siri or alexa, which can assist users in solving real-life data science problems.

When: Wednesday, 12th August, 2020 (9:00 PM)
Where: Online. Zoom Meeting ID: 682 8963 2557, Password: 162219
Organized by BUET ACM Chapter, Dept. of CSE, BUET.