Shubhra Kanti Karmaker "Santu"

Email: sks0086@auburn.edu
Portfolio: https://karmake2.github.io/
Mobile: +1-217-979-3244
Google Scholar: https://scholar.google.com/citations?user=y6pZKT4AAAAJ

Current Position

- **Auburn University** — Auburn, Alabama, US
  Assistant Professor — *Jan 2020 - Present*
  Department of Computer Science and Software Engineering

Research Area

- Big Data, Information Retrieval, Natural Language Processing, Machine Learning

Academic Preparation

- **Massachusetts Institute of Technology (MIT)** — Boston, Massachusetts, US
  Postdoctoral Research Associate — *Jan 2019 - Dec 2019*
  Laboratory for Information & Decision Systems (LIDS)
  Host: Kalyan Veeramachaneni

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Ph.D. in Computer Science — *Aug 2014 - Dec 2018*
  Advisor: ChengXiang Zhai
  CGPA: **4.00** out of 4.00

- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  M.S. in Computer Science & Engineering — *Apr 2012 - Apr 2014*
  CGPA: **4.00** out of 4.00

- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  B.S. in Computer Science & Engineering — *Jun 2007 - Feb 2012*
  CGPA: **3.98** out of 4.00. [**Class Rank: 1** (out of 126 in the entire department)]

Grants

- [**Awarded**] Deep Time Series Forecasting Techniques for modeling Molecular Dynamics
  - **Principal Investigator**
  - Sponsor: Arizona State University
  - Amount: **$23,135**
- [Under Review] IES proposal with budget of ∼$900,000 as PI
- [Under Review] Two NSF CISE CORE small proposals each with a budget of ∼$500,000 as PI
- [Under Review] NSF Research on Emerging Technologies for Teaching and Learning (RETTL) with a budget of ∼$700,000 as PI

Selected Services

- **Panelist**: NSF/CISE Grant Proposal Review Panel, 2020
- **PC Member**: **SIGIR**[2020, 2021], **ACL**[2019,2021], **WSDM**[2021, 2020, 2019], **IUI**[2021, 2020, 2019]
- **Reviewer**: IEEE Transactions on Knowledge and Data Engineering (TKDE)
- **Reviewer**: Knowledge and Information Systems Journal, Neuro-Computing Journal
- **PC Member**: WSDM Demo Track [2019, 2018], IUI Demo Track [2018,2019]

Previous Industrial Experience

- **Microsoft Research** — Bellevue, Seattle, US
  Summer Research Intern @FUSE labs — *Summer, 2018*
  Project: Developed semantic representations for understanding utterances for LUIS
  Mentor: Riham Mansour

- **Microsoft Research** — Redmond, Seattle, US
  Summer Research Intern @Internet Service Research Center (ISRC) — *Summer, 2017*
  Project: Automatic Self-Evolving Text Generation
  Mentor: Hao Ma

- **Yahoo Research** — Sunnyvale, California, US
  Summer Research Intern @Search Science Group — *Summer, 2016*
  Project: Influence Modeling for User Search Behavior
  Mentor: Yi Chang

- **@WalmartLabs** — Sunnyvale, California, US
  Summer Research Intern @Search Relevance Team — *Summer, 2015*
  Project: Learning-to-Rank for E-Commerce Search
  Mentor: Parikshit Sondhi

- **Jump Trading LLC** — Chicago, Illinois, US
  Text Mining Consultant @Search Relevance Team — *Fall 2016 - Spring 2017*
  Project: *Confidential and Restricted*

- **Stochastic Logic, ACI Limited** — Dhaka, Bangladesh
  Quantitative Software developer — *May 2012 - August 2014*
  Worked on stock market forecasting using time series analysis

Teaching Experience

- **Auburn University** — Dhaka, Bangladesh
  Assistant Professor, Department of CSSE — *Jan 2020 - Present*
  Courses Taught: Information Retrieval (Senior Undergrad Level), NLP (Graduate Level).

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Instructor, Department of Computer Science — *Fall 2018*
  Courses Taught: Text Mining Seminar (Graduate Level).

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Teaching Assistant, Department of Computer Science — *Fall 2017*
  Courses Taught: Advanced Information Retrieval (Graduate Level).

- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  Lecturer, Department of Computer Science & Engineering — *May 2012 - August 2014*
  Courses Taught: Computer Graphics, Computer Architecture, Operating Systems etc.

Publications

- [**IUI 2021**] : Biddut Sarker Bijoy,Syeda Jannatus Saba, Souvika Sarkar, Md Saiful Islam, Sheikh Rabiul Islam, Md. Ruhul Amin, **Shubhra K Karmaker**. "*COVID19α: Interactive Spatio-Temporal Visualization of COVID-19 Symptoms through Tweet Analysis*". In ACM IUI, 2021.
- [**CIKM 2020 (**Best Poster Nomination**)**] : **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi and ChengXiang Zhai. "*Empirical Analysis of Impact of Query-Specific Customization of nDCG: A Case-Study with Learning-to-Rank Methods*". In ACM CIKM, 2020.
- [**ICWSM 2020**]: Naeemul Hassan, Amrit Poudel, Jason Hale, Claire Hubacek, Khandakar Tasnim Huq, **Shubhra Kanti Karmaker Santu**, Syed Ishtiaque Ahmed. "*Towards Automated Sexual Violence Report Tracking*". In ICWSM 2020.

- [**CoNLL 2019**]: **Shubhra Kanti Karmaker Santu**, Kalyan Veeramachaneni, ChengXiang Zhai. "Neural Language Models with Evolving Topical Influence". In ACL SIGNLL CoNLL 2019.
- [**CIKM 2019**]: Saar Kuzi, Sahiti Labhishetty, **Shubhra Kanti Karmaker Santu**, Prasad Pradip Joshi and ChengXiang Zhai. "Analysis of Adaptive Training for Learning to Rank in Information Retrieval". In ACM CIKM 2019.
- [**SIGKDD Explorations 2018**]: **Shubhra Kanti Karmaker Santu**, C. Geigle, D. C. Ferguson, W. Cope, M. Kalantzis, D. Searsmith, Chengxiang Zhai. "SOFSAT: Towards a Setlike Operator based Framework for Semantic Analysis of Text". In ACM SIGKDD Explorations [Position Paper] 2018.
- [**CIKM 2018**]: **Shubhra Kanti Karmaker Santu**, Liangda Li, Yi Chang, ChengXiang Zhai. "JIM: Joint Influence Modeling for Collective Search Behavior". In ACM CIKM 2018.
- [**WPES 2018**]: **Shubhra Kanti Karmaker Santu**, Vincent Bindschaedler, ChengXiang Zhai, Carl A. Gunter. "NRF: A Naive Re-identification Framework". In WPES@ACM CCS 2018.
- [**CIKM 2017**]: Yiren Wang, Dominic Seyler, **Shubhra Kanti Karmaker Santu**, ChengXiang Zhai. "A Study of Feature Construction for Text-based Forecasting of Time Series Variables". In ACM CIKM [Short Paper] 2017.
- [**SIGIR 2017**]: **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi, ChengXiang Zhai . "On Application of Learning to Rank for E-Commerce Search". In ACM SIGIR 2017.
- [**WWW 2017**]: **Shubhra Kanti Karmaker Santu**, Liangda Li, Dae Hoon Park, Yi Chang, ChengXiang Zhai . "Modeling the Influence of Popular Trending Events on User Search Behavior". In WWW 2017.
- [**CIKM 2016**]: **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi, ChengXiang Zhai . "Generative Feature Language Models for Mining Implicit Features from Customer Reviews". In ACM CIKM 2016.
- [**IJCNN 2014**]: Md. Mustafizur Rahman, **Shubhra Kanti Karmaker Santu**, Md. Monirul Islam, Kazuyuki Murase. "Forecasting time series - A layered ensemble architecture". In IJCNN 2014.
- [**CEC 2014**]: **Shubhra Kanti Karmaker Santu**, Md. Mustafizur Rahman, Md. Monirul Islam, Kazuyuki Murase . "Towards better generalization in Pittsburgh learning classifier systems". In IEEE CEC (Congress on Evolutionary Computation) 2014.

Ongoing Research Projects

- **TextMall: Text-Mining Made Simple for All**: The objective of this research is to develop a general-purpose text analytics platform, i.e., Text-Mall, which would enable real-world users to easily explore the power of Text-Mining in a simple and interactive fashion without worrying about the underlying details of Natural Language Processing. This project aims to address this challenge by developing a new general-purpose text analytics platform, i.e., Text-Mall (aka. Text Mining for all), which will provide simple intuitive operators for interactive text-mining tasks which can be easily explained and taught to the general public.
- **VIDS: Virtual Interactive Data Scientist**: What if an AI agent can serve you as your personal data scientist? The goal of this project is to develop VIDS (Virtual Interactive Data Scientist), which can be thought of as an intelligent agent that, given a large corpus of data, starts talking to the domain experts like a human data scientist in order to quickly figure out the user's need and explore alternative prediction opportunities. It will automatically formulate different prediction tasks, as well as explore alternative learning models to recommend the best one to the users through natural conversation.
- **Annotate: Concept Annotation from Users Perspective**: Information retrieval and Knowledge mining become much easier if data is categorized and annotated precisely. With the rapid growth of Big-data, it is infeasible to perform manual annotation, as it is slow and expensive. Although the area of text annotation is not in the nascent phase, it has not been well-studied from a user-centric point of view, which is the goal of this project.
- **Robust IR Evaluation**: Previous studies have shown that popular Information Retrieval (IR) evaluation metrics, e.g., $nDCG$, ERR, MAP, are not robust with respect to the variation of the cutoff rank k. In this project, our main goal is to investigate how we can make IR evaluation metrics more robust; more specifically, less sensitive to the variation of cutoff k.

- **A2I-MOOC**: MOOCs have abysmal retention rates (5-15%) and high student failure rates (7-13%). In this project, we propose two ways to increase student engagement in MOOCs and other online courses through an artificially intelligent system that leverages machine learning and natural language processing. This system will (1) process, prioritize and organize students' questions in real-time and provide the most relevant questions to instructors for answering during their live lectures, and (2) automate the creation of breakout rooms (which have recently become popular in Zoom classes) based on high-interest topics emerging from student questions and populated by like-minded students during live lectures.
- **Molecular Dynamics Modeling with Deep Time series Forecasting**: In this project, We pare studying the state of the art machine learning algorithms to model the outcome of classical Molecular Dynamics (MD) simulation. By deploying these algorithms, we expected to bypass the need to perform numerical integration over millions of dimensions, and yet predict the outcomes of long-timescale MD simulations. The algorithms, including Fully connected Feed-forward Neural Networks, Recurrent Neural Networks with LSTM / GRU memory cells, will be tested to predict conformational transitions on a range of biological systems starting from simple proteins up to complex multimers.

Students

- 7 Ph.D. Students
    - Alex "Ralph" Knipper [2020 - Current]
    - Md. Mahadi Hasan "Sibat" [2020 - Current]
    - Souvika Sarkar [2020 - Current] (Under-Represented)
    - Naman Bansal [2020 - Current]
    - Dongji Feng [2020 - Current]
    - Mousumi Akter [2020 - Current] (Under-Represented)
    - Minh Smith [2020 - Current]
- 1 Undergraduate Student
    - Steven Woolf (Spring 2021)

Honors and Awards

- Best Poster Award Nomination at CIKM 2020
- SIGIR Travel Grant for attending CIKM 2018
- Dean's list award (For brilliant result in each year during undergraduate program at BUET)
- Crest of honor (Highest CGPA in the department, presented by BUET alumni association)