Kanchan Chowdhury

Email : kchowdh1@asu.edu
Webpage: kanchanchy.github.io
Mobile : +1-480-410-8677
Linkedin: linkedin.com/in/kanchan-chowdhury-5729699a

Research Interests

Machine Learning, Database Systems, Big Data Analytics, and Geospatial Data Analytics

Technical Skills

- **Languages & Databases**: Python, Java, C, Scala, SQL, MySQL, PostgreSQL, and SparkSQL
- **Others**: Apache Spark, Apache Sedona, GeoPandas, PyTorch, Scikit-learn, Spatial and Statistical Data Analysis, Machine Learning and Deep Learning, Pandas, Numpy, Git, Rest API, Android SDK, IntelliJ IDEA, Eclipse, and Data Visualization with Matplotlib, Seaborn, and Plotly Express

Education

- **Arizona State University** — Tempe, Arizona
  PhD in Computer Science — *Aug. 2018 – Present*
  Advisor: Prof. Mohamed Sarwat, CGPA: 4.00
- **Chittagong University of Engineering and Technology** — Chittagong, Bangladesh
  Bachelor of Science in Computer Science and Engineering — *Mar. 2010 – Nov. 2014*
  Advisor: Prof. Mohammed Moshiul Hoque, CGPA: 3.76

Experience

- **Arizona State University** — Tempe, Arizona
  Research & Teaching Assistant — *Aug. 2018 - Present*
  - **Research**: 1) Solving research problems in the intersection of spatio-temporal big data analytics and machine learning to enable efficient application of scalable spatial data into model training. 2) Designing and implementing a spatiotemporal deep learning framework. 3) Evaluating state-of-the-art approaches for synthesis of SQL queries from natural language questions with four real-world datasets.
  - **Instructor of the Course**: Spatial Data Science and Engineering (Fall 2022)
  - **Courses Taught as TA**: Distributed Database Systems, Data Processing at Scale, Object-Oriented Programming & Data Structure, Principles of Programming with C++, Principles of Programming with Java & Python
- **Gagagugu PTE LTD** — Dhaka, Bangladesh
  Software Engineer — *Jan. 2017 - Jun. 2018*
  - **Responsibilities**: Developing Android Apps with social networking features such as calling, messaging, and posting.
- **Le Chef Plc** — Dhaka, Bangladesh
  Android Application Developer — *Jan. 2015 - Dec. 2016*
  - **Responsibilities**: Developing Android Apps featuring online order and reservation services for restaurants in UK.

Publications

- **Kanchan Chowdhury**, Mohamed Sarwat; GeoTorch: A Spatiotemporal Deep Learning Framework. *30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, 2022
- **Kanchan Chowdhury**, Vamsi Meduri, Mohamed Sarwat; A Machine Learning-Aware Data Re-partitioning Framework for Spatial Datasets. *38th International Conference on Data Engineering (ICDE)*, 2022
- Vamsi Meduri, **Kanchan Chowdhury**, Mohamed Sarwat; Evaluation of Machine Learning Algorithms in Predicting the Next SQL Query From the Future. *ACM Transactions on Database Systems (TODS)*, 2021
- Jia Yu, **Kanchan Chowdhury**, Mohamed Sarwat; Tabula in Action: A Sampling Middleware for Interactive Geospatial Visualization dashboards. *46th International Conference on Very Large Databases (VLDB)*, 2020.

- Vamsi Meduri, **Kanchan Chowdhury**, Mohamed Sarwat; Recurrent Neural Networks for Dynamic User Intent Prediction in Human-Database Interaction. *22nd International Conference on EDBT, 2019*
- **Kanchan Chowdhury**, Lamia Alam, Shyla Sarmin, Safayet Arefin, Mohammed Moshiul Hoque; A Fuzzy Features Based Online Handwritten Bangla Word Recognition Framework. *18th ICCIT*, 2015

Projects

- **Named Entity Recognition**: This work tunes various steps of state-of-the-art methods for named entity recognition in order to experiment the changes in performance. Evaluation is done with two popular datasets: CoNLL-2003 and OntoNotes-5.0.
- **Spatial Data Repartitioning**: A framework built on Apache Sedona for efficient re-partitioning of spatial data with an end objective to reducing training time and memory usage of spatial machine learning models. It can reduce the training time significantly without major impact on model accuracy.
- **Climate Change Forecasting**: A data science project to perform data cleaning, feature engineering, and data preprocessing operations on raw temperature data and predict temperature trend with LSTM model.
- **Data Visualization**: A project to visualize statistical data in Python with the help of most frequently used plots required for data analysis. It uses three libraries: Matplotlib, Seaborn, and Plotly Express.
- **NLIDB-Bench**: A benchmark for evaluating state-of-the-art approaches of SQL query generation from natural language text queries. Besides proposing a set of evaluation metrics, we evaluate all approaches with three state-of-the-art datasets along with our own proposed dataset.
- **Fake News Detection**: A data science project to perform data cleaning, feature engineering, and data preprocessing operations on news datasets and classifying fake and real news with Bidirectional LSTM model.
- **Sentiment Analysis with BERT**: A project to classify twitter emotions with BERT pretrained model.
- **Hotspot Analysis on Apache Sedona**: This work performs spatial queries and range joins between two spatial datasets and calculates Getis-Ord statistic of NYC Taxi Trip dataset to perform hot-cell analysis.
- **Image Denoising with Autoencoder**: This project adds noise to MNIST dataset images and builds an auto-encoder model consisting of encoder and decoder to reproduce actual images.
- **Optimizing Hyperparameters of Deep Learning Model**: Comparing approaches for searching the optimum hyper-parameters of a deep learning model, such as grid search and Bayesian optimization.

Participation and Awards

- Recipient of CIDSE Doctoral Fellowship at Arizona State University for the academic year 2018-2019.
- Contributed to Apache Sedona, an open-source geospatial cluster computing framework, by adding supports for two new spatial data types. I am Serving as a PPMC member and release manager of Apache Sedona.
- 2nd Runner-up at National Hackathon organized by ICT Division of Bangladesh in 2014. The challenge of the hackathon was to design a project based solution to solve a national problem of the country.
- 2nd Runner-up at National Mobile Application Code Hub organized by BUET, Bangladesh in 2014.
- 6th at Inter University Programming Contest organized by CUET, Bangladesh in 2012.
- Recipient of Honors award from undergraduate University for maintaining good CGPA.

CourseWorks

Distributed Database Systems, Statistical Machine Learning, Fundamentals of Statistical Learning, Data mining, Semantic Web Mining, Programming in C/C++, Object Oriented programming, Data Structure, Algorithms, Discrete Mathematics, Artificial Intelligence, Software Engineering, Operating Systems, Database Management Systems, Social Media Mining, Big Data Analysis with Scala and Spark, Deep Neural Networks with PyTorch, and Effective Programming in Scala.