

EDUCATION

Khoury College of Computer Sciences , Northeastern University

Seattle, WA

- M.S. in Computer Science. –GPA 3.89

Expected 12/2022

College of Science & Engineering , University of Minnesota - Twin Cities

Minneapolis, MN

- M.S. in Industrial Engineering - Data Analytic track. –GPA 3.51
- B.S. in Mathematics & Statistics. –GPA 3.77 with Distinction
- 2018 UMN Industrial and Systems Engineering Merit Fellowship

12/2020

05/2018

SKILLS & COURSEWORK

Programming : Access, AWS, Azure, Apache Atlas, AMPL, C, C++, CSS, DVC, Docker, CPLEX, EXCEL, Google Cloud (Big Query), Hive, Hackolade, HTML, Hadoop, Java, JavaScript, LONDO, MongoDB, Mathematica, MATLAB, Python (PyTorch, TensorFlow, spaCy, Dask), PowerPoint, R, React, React Native, SAS, SQL, Tableau, Word

Computer Course Emphasis: Algorithm Design, Design Pattern, Software Engineering, Cloud Computing, Machine Learning, Database, Distributed System, Web Development [DEMO](#)

Portfolio: zitaoshen.rbind.io [Click here for website](#)

PUBLICATION

- Zitao Shen*, Dalton Schutte*, Yoonkwon Yi*, Anusha Bompelli, Fang Yu, Yanshan and Rui Zhang. 2021. Classifying the lifestyle status for Alzheimer's disease from clinical notes using deep learning with weak supervision. BMC Medical Informatics and Decision Making.
- Ruoyan Kong, Ruobing Wang, Zitao Shen. 2021. Virtual Reality System for Invasive Therapy. 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)

WORK EXPERIENCE

Software Development Engineer Intern

05/2022–08/2022

Amazon AWS

Seattle, WA

- Working in an AWS EC2 team using scrum development methodology.
- Designing a reliable system and breaking it down iteratively to meet customers' needs
- Using Ruby on Rails to set up a scalable and automatized backend service modules to replace some manual process
- Modifying the front-end UI to ensure users have a good experience with new features.

Machine Learning Engineer Intern

05/2021–08/2021

Seagate

Remote

- Wrote a Python data cleaning script for multiprocessing million of internal text documents.
- Set up a docker container for pre-training a large BERT-based language model with Data Vision Control (DVC), PyTorch, and CUDA.
- Designed, built, and deployed a hard drive reliability control machine learning pipeline in a KNIME server for real-time factory production. The evaluation metric was improved by 22% by comparing it with the baseline model.
- Built a LightGBM model pipeline for real-time hard drive quality control, and the recall reached 0.93 on test data.

Data Modeling Engineer Intern

10/2020–12/2020

MilliporeSigma

Remote

- Wrote a python script for collecting, generating metadata information into configuration files to improve the internal data category
- Constructed several data modeling schemas from Hive by using Hackolade and Python
- Using Python package Dask to implement scalable parallelized machine learning models and wrote internal Dask tutorial documents.

Nature Language Processing Research Assistant

05/2020–12/2020

Zhang's Health Informatics Research Lab

University of Minnesota

- Using Natural Language Processing with Deep Learning method to label two million jobs, in term of skill types, job complexity and value chain phases, based on the related job descriptions.
- Designing annotation pipeline and implementing automated annotation tasks on Amazon Mturks with AWS service
- Building deep learning models, Bi-LSTM and BERT, to extract lifestyle exposures of Alzheimer patients from textual clinical records

PROJECTS

Wells Fargo Campus Analytics Challenge 2020 (Awarded with Finalist+ Grand Prize) [Click here for website](#)

07/2020–08/2020

- Proposed a novel hybrid classification model (Neural Networks + LightGBM) to classify imbalanced binary labels
- This mode had an outstanding performance with averaged F1 score-0.96 on the testing data among hundreds run.
- The project was awarded with Finalist (8 out of 28) + Grand Prize (4 out of 28).

Ethereum Loan Market Study

12/2019–04/2020

- Volunteered in a cross-functional team in an Agile environment for a Fintech startup, DeFiner.
- Deployed a data pipeline for collecting Twitter user's data on Ethereum with AWS Lambda and S3, and implemented a sentimental analysis to relate with the fluctuation of the Ethereum loan market in Python. The R-squared reached 83%.
- Set up a Tableau's dashboard to visualize and track critical indicators of Ethereum loans' markets, such as borrowing, loan, and liquidation. The dashboard was praised and adopted as a showcase inside the company