Zitao Shen

1046 29th Ave SE
Minneapolis, MN 55455
Phone: (949)829-2518 E-mail: shenx499@umn.edu
Website: zitaoshen.rbind.io

Education

Khoury College of Computer Sciences , Northeastern University Seattle,WA

- M.S. in Computer Science Expected 12/2022

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

- M.S. in Industrial Engineering - Data Analytic track 12/2020
  - — GPA 3.51
  - — 2018 UMN Industrial and Systems Engineering Merit Fellowship

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

- B.S. in Mathematics, double majored in Statistics 05/2018
  - — GPA 3.77, with Distinction

Coursework

Computer Course Emphasis: Algorithm Design, Data Mining, Reinforcement Learning , Machine Learning, Natural Language Processing, Database

Mathematics Course Emphasis: Optimization, Stochastic Modeling(emphasized on Markov Decision Process), Statistical Learning, Nonparametric Statistic Methods

Operation Research Course Emphasis: Dynamic Inventory Control, ICU Discharge Policy Design, Lean Process

Portfolio: zitaoshen.rbind.io *Click here for website*

Skills

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

Work & Internship

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, evaluation metric, 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 machine learning and deep learning models, such as Bi-LSTM and BERT, to extract lifestyle exposures of Alzheimer patients from textual clinical records

Graduate Research Assistant *03/2018–09/2019*

Department of Industrial and Systems Engineering *University of Minnesota*

1. Topic: Promotion strategy for a service firm with delay sensitive customers. *Click here for website*.
   - Gathered, cleaned and organized large data file with over 100,000 restaurants' service information from Yelp's review and Google Map by using R and SQL
   - Quantified and visualized Yelps' customers' review by text mining in R
   - Built a Two-Stage Least Square Regression models to verify the proposed Queueing Model. The R-squared reached 79%.
2. Topic: Queueing model's application on blockchain transaction
   - Collected, cleaned and visualized bitcoin historical network activity and mempools' daily transactions in SQL
   - Built a rule-based classifier to distinguish bitcoin historical events by using R. The prediction reached 87% accuracy

Program Evaluator *09/2017–03/2018*

SMART Learning Commons *University of Minnesota*

- Suggested and led the first-time formal evaluation on measuring the success of SMART Learning Program by studying and analyzing survey data in R. *Click here for website*
- Built a predictive model (ordinary logistic regression) to predict daily students' length of visiting based on interested KPIs
- As result, this evaluation's results has been integrated into mentor training, by emphasizing the importance of the length of service time

Statistical Consultant Volunteer *02/2017–03/2017*

Undergraduate Statistics Club & Project for Pride in Living *University of Minnesota*

- Analyzed data to study current residents' housing preference for a local NGO, Project for Pride in Living, based on their offered housing types, location, and other elements in Excel and R
- Cooperated with a cross-functional team and hosted weekly meeting between the team and the organization

PROJECTS

Wells Fargo Campus Analytics Challenge 2020 (Awarded with Finalist + Grand Prize) *Click here for website* *07/2020*

- 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*

- Worked 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

Improving the Repositioning Operation for Bike Rental System *Click here for website* *03/2019–05/2019*

- Designed and compared dock-based bikes' and dock-less' inventory rebalancing strategy under a large-scale service network, in the term of a potential mismatch between rental/return demand
- Formulated the problem as a Markov Decision Process model based on the parameters generated from a local public bike company's service record in R
- Implemented a simulation on checking proposed policy's performance in MATLAB. The proposed policy reduced the percentage of mismatch by 32%

Optimizing Intensive Care Unit Discharge Decisions Policy for Burn Units *Click here for website* *03/2019–05/2019*

Project for Stochastic Dynamic Optimization *University of Minnesota*

- Designed demand-driven ICU discharge policies for burn units by describing the process as a Markov Decision Process
- Compared with several proposed discharging polices, included ranking patients based on different medical cost metrics, by implementing related simulation studies in Python

PUBLICATION

- Zitao Shen*, Yoonkwon Yi*, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 2020. Natural Language Processing Methods to Extract Lifestyle Exposures for Alzheimer's Disease from Clinical Notes. HealthNLP 2020 (*first author)
- Ruoyan Kong, Ruobing Wang, Zitao Shen. 2021. Virtual Reality System for Invasive Therapy .2021 IEEE Conference on Virtual R
- Zitao Shen*,Dalton Schutte*, Yoonkwon Yi*, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 2021. Classifying the Lifestyle Status for Alzheimer's Disease from Clinical Notes Using Deep Learning with Weak Supervision.BMC Medical Informatic

Teaching Experience

Graduate Teaching Assistant *09/2018–05/2020*

Department of Industrial and Systems Engineering *University of Minnesota*

- Assisted for graduate-level IE lean process and stochastic modeling courses, including mentoring, lecturing, grading, and clerical help
- Host and led weekly office hour and end-of-semester review sessions
- Used R markdown and Latex to design weekly homework solutions and R tutorial handouts

Peer Learning Consultant *09/2017–05/2018*

SMART Learning Commons *University of Minnesota*

- Tutored college students about college-level math and statistics for 5 hours per week
- Used Mathematica, R studio and IntelliJ IDEA to answer students' questions, and teach them how to use the software

Leadership & Volunteer Work

International Student Consultant *06/2018-08/2018*

College of Science & Engineering *University of Minnesota*

- Offered consulting service about the college life to new international students during the college orientation
- Set up two new online-based surveys to collect students' feedback on their orientation experience

International Buddy Program Mentor *08/2016-12/2016*

International Student and Scholar Services Global Programs *University of Minnesota*

- Helped new International students to adapt to the new environment quickly
- Set up a daily communication channel via emails and phone calls to help to resolve students' issues remotely with effectiveness