Mingyang (Mia) Wan

Email: w1996@tamu.edu miaapplyjob@gmail.com
Phone: (979) 224-7381 Github: https://github.com/mia1996
Address: 514C H.R. Bright Building, College Station, TX 77843-3112

EDUCATION:

Texas A&M University, College Station, TX Expected graduate in 09/2022
- Master of Computer Science, GPA: 4.0/4.0
- Coursework: Database, Algorithms, AI, Web Development, Operating System, Software Engineering

Texas A&M University, College Station, TX College Station, TX, 12/2020
- Master of Science in Industrial Engineering (Data Science)
- Coursework: Data Analysis, Dynamic Programming, Database

Beijing University of Posts and Telecommunications Beijing, China, 04/2018
- Bachelor of Engineering in Logistics (Automation)

HIGHLIGHTS:

I'm an engineer and researcher working at Data Lab of Texas A&M University, focusing on Anomaly Detection and Fairness AI. I am also a seasoned software developer working at Texas A&M Data Lab, contributed to several Python open-source projects with more than **1700** GitHub stars and more than **50,000** downloads in total.

TECHNICAL SKILLS:

Programming Language: Python, Java, C/C++, Matlab, Shell Script, HTML, SQL, Spark, Ruby
Framework / Tools: AWS S3, Git, Linux, Bootstrap, Tableau, Tensorflow, PyTorch, Scikit-Learn, MySQL, Azure Blob Storage, Google Cloud Storage

PROFESSIONAL EXPERIENCE:

Research Assistant at DATA lab at Texas A&M University (College Station, TX) ***05/2020-Now***
- Developing time-series anomaly detection algorithms for an open-source project (TODS project).
- Implemented data pipelines and test cases to build and maintain essential primitives. Contributed to the team with preprocessing and anomaly detection modules (such as AutoEncoder, VAE, etc.).
- Developed key metrics to track the performance of fraud and verification strategies in Meta-AAD (Active Anomaly Detection with Deep Reinforcement Learning).

Data Scientist, Co-Op at Bristol-Myers Squibb (Devens, MA) ***07/2019-12/2019***
- Collaborated with multi-functional teams to maintain data pipelines and back-end services for real-time decisions in Capacity Model project by designing and validating logic queries using SQL and Python; saved the organization over $30,000 and achieved Yellow Belt certificate.
- Performed data extraction and manipulation from database; implemented case-oriented visualization solutions by Tableau for biopharma business analytics across the product lifecycle, reaching 80% improvement of efficiency on weekly reports.
- Exploited rich datasets in AS project by using open-source libraries with Git and AWS. Researched on NLP to improve ongoing analysis processes, simplifying support for QC to filter delay status every 12 hours.

PROJECT WORK:

TODS: An Automated Time-series Outlier Detection System (> 400 stars) ***03/2021***
- Contributed to the Anomaly Detection project: https://github.com/datamllab/tods
- Participated in the full development cycle: design, develop, experiment, analyze, and deploy. Mainly devoted to preprocessing and anomaly detection modules (including AutoEncoder, VAE, etc).

RAID-on-Cloud: A Cloud-based Replication NAS Backed by Multiple Providers ***10/2020***

- Developed a client-side application and access multiple cloud storage services (AWS S3, Azure Blob Storage, and Google Cloud Storage) via APIs and SDKs.
- Balanced the utilization of cloud resources across platforms and implemented the traditional, POSIX-style filesystem functions, such as read() and write(), as wrappers of object storage APIs, and to implement a RAID-like replication system.

PySpark Programming: Intrusion Detection Using Big-Data Analysis ***09/2020***

- Parallel Computing through PySpark to implement the algorithms for detecting intrusive behaviors inside a network log.
- Program the Spark interfaces to perform big-data analysis on statistical data with RDD.

Developed Toolkit for Reinforcement Learning in Card Games (>1300 stars) ***04/2020***

- Contributed to the Reinforcement Learning (RL) in card games: https://github.com/datamllab/rlcard
- An official tutorial for RLCard. Provided step-by-step instructions and running examples for both Python and R. View this project on https://github.com/datamllab/rlcard-tutorial.

Developing progress monitor System for University Writing Center ***01/2020***

- Designed user interface using JavaScript, jQuery, Bootstrap, HTML and CSS. Utilized the Agile development life cycle, managed projects in Pivotal Tracker with user stories and customer issues. This app is deployed on https://shrouded-chamber-95563.herokuapp.com/.

Developing models to predict the risk of real case insurance claim ***10/2019***

- Preprocessed dataset containing 10,000 policies and built data-mining models predicting insured cars; compared feasible models and found out the optimal prediction solution.
- Evaluated Tree-based models on the test set and achieved 84% accuracy improvement compared with the previous model.

PUBLICATIONS:

Manuscript — **Mingyang Wan**, Daochen Zha, Ninghao Liu, Na Zou, Xia Hu
Techniques for Fairness-Aware Machine Learning
- Survey on In-processing Machine Learning techniques for fairness models and organize approaches into the novel framework of explicit and implicit mitigation methods, subcategorizing into a further 10 method areas.

ICDM'20 — Daochen Zha, Kwei-Herng Lai, **Mingyang Wan**, and Xia Hu. Meta-AAD: Active Anomaly Detection with Deep Reinforcement Learning. IEEE International Conference on Data Mining, 2020.
- Contributed to the development of strategies, rules and the execution used in predicting and preventing fraud.

AAAI'21 — Kwei-Herng Lai*, Daochen Zha*, Guanchu Wang, Junjie Xu, Yue Zhao, Devesh Kumar, Yile Chen, Purav Zumkhawaka, **Mingyang Wan**, Diego Martinez, Xia Hu.
TODS: An Automated Time Series Outlier Detection System. AAAI Conference on Artificial Intelligence, demo track, 2021.
- Participated in the full development cycle: design, develop, experiment, analyze, and deploy. Mainly devoted to preprocessing and anomaly detection modules (including AutoEncoder, VAE, etc).