ANNA DAI

www.linkedin.com/in/dai-anna ✉ anna.dai@duke.edu github.com/dai-anna

EDUCATION AND CERTIFICATIONS

Duke University   **Durham, NC**
Master in Interdisciplinary Data Science (MIDS)   08/2021 – Present
- **Duke Cloud Club Co-Founder and President**: Created AWS Organization for MIDS students and managed funds for members
- Past Coursework: Machine Learning, Cloud Computing, NLP, Mathematical and Inferential Statistics, Causal Inference
- Current Coursework: Deep Learning, Deep Reinforcement, Data Structures and Algorithms, Data Visualizations

University of California, Los Angeles   **Los Angeles, CA**
Bachelor of Arts, Business Economics   09/2012 – 08/2016

AWS Certified Solutions Architect - Associate
- Demonstrated proficiency in designing, implementing, and deploying complex distributed systems on AWS

Certified Public Accountant in the State of California

WORK EXPERIENCE

Duke University | The Graduate School   **Durham, NC**
Teaching Assistant   08/2022 – Present
- Held weekly office hours to explain course concepts, answer questions and graded student projects for Data Engineering Systems

Stitch Fix | Merchandising Algorithms   **Remote / San Francisco, CA**
Data Science Intern   06/2022 – 08/2022
- Researched unexplored problem space of incorporating external fashion trends into currently historical-data-driven algorithms
- Developed steel-thread attempt at mapping trending terms to inventory items by customizing internal search pipeline and leveraging pre-trained word embeddings and latent item embeddings to enable trend-level modeling
- Discovered new data source from existing vendor and automated data cleaning and term extraction pipeline from PDF reports to build a database of high-quality trending terms by date and category (i.e. Womens, Mens, Children, etc.)

Duke University | Kenan Institute for Ethics Purpose Project   **Durham, NC**
Race and the Professions Fellow   09/2021 – 05/2022
- Represented and voiced the Asian (Canadian) perspective at meetings to address racial justice issues within the professions
- Organized data visualization contest to spotlight racial inequality and racism within various Duke University departments

Ernst & Young LLP | Financial Services Office   **San Francisco, CA**
Senior Tax Consultant   09/2018 – 04/2021
- Led US tax compliance and consulting teams of five to ten staffs as acting manager, reviewed staff and senior deliverables, and juggled up to 14 engagements at any given time from clients who managed up to $33 billion AUM
- Conducted research tax credit studies, cleaned, and manipulated raw personnel data to quantify credit values, interviewed clients' technical personnel, researched software engineering and data science topics, and drafted technical memoranda

Tax Staff   01/2017 – 09/2018
- Wrote advanced excel-based data validation logic to reduce human error in performing trading security analysis and preparing tax allocations, thus improving reviewability, updatability and reproducibility for future years
- Assigned specifically to handle challenging entities involving multi-State income sourcing and apportionment, numerous side-pocket investments, debt-backed funds, and large management companies picking up from over 15 underlying funds

PROJECT EXPERIENCE

Duke Anesthesiology Drug Diversion Intervention | Capstone project   **Durham, NC**
- Partnered with Duke Anesthesiology and Fuqua professors to develop ensemble learning and unsupervised anomaly detection models for detecting with high sensitivity the early phases of drug diversion by anesthesiologists
- Researched and experimented with algorithms to overcome extreme data imbalance and noise from varied procedural practices

Automatic Tweet Generator | AWS Cloud-Based Data Engineering project   **Durham, NC**
- Built a cloud-native data-scraping and processing application that automatically generates Tweets for current trending hashtags using a Hidden Markov Model based on N-Grams in Python
- Leveraged AWS Lambda, Python CDK, EventBridge, Batch, EC2 Spot Instance, and S3 services as well as Docker containers

Deep Document Classification | NLP Benchmarking Project   **Durham, NC**
- Collected Tweets on 7 chosen topics and let LDA infer latent document-topics and generate synthetic data on inferred topics
- Benchmarked LDA and bi-directional GRU neural network with self-trained word embeddings on classifying documents into topics on synthetic and real data – implemented in Python Keras TensorFlow

SKILLS

- Programming Languages and Frameworks: Python, R, and SQL; basic C++
- Tools and Services: Visual Studio Code, R Studio, Git, Docker, FastAPI, Amazon Web Services, and Google Cloud Platform
- Machine Learning (PyTorch, TensorFlow, Scikit-Learn) and Statistical Modeling (linear and logistic regressions, time series)
- Natural Languages: English, Chinese (Mandarin), and French; conversational Korean