Naomi (Yu Wen) Chiu

ychiu60@gatech.edu | linkedin.com/in/naomichiu | https://github.com/chiuyuwen91/Portfolio | 415-691-9536

Highly motivated engineer with passion for data engineering. Managed and built data piplines Airflow (DAG) in Spark on AWS EMR cluster. Great working knowledge about data warehouse concepts, distributed system, machine learning, and analytics. Proficiency in Python, SQL.

TECHNICAL SKILLS

- **Programming Languages**: SQL, Python, R, Java
- **Visualization Tools**: ArcGIS, Tableau
- **Databases Systems**: DBT, AWS (Redshift, S3)
- **Web Development**: Django, Docker, REST APIs
- **Project Management**: Agile Scrum, Git, JIRA
- **Machine Learning (ML)**: Deep Learning, Natural Language Processing, Computer Vision, PyTorch, Tensorflow, Scikit-learn
- **Data Streaming**: Kafka, Spark
- **Statistical Modeling**: Experimental Design, A/B Testing, Hypothesis Testing, Causal Inference

EXPERIENCES

University of San Francisco, ***Teaching Assistant***, San Francisco, CA **Aug 2021 – Dec 2022**

- Taught courses: **Microeconomics**, **Experiments & Causal Inference**, and **Advanced Experimentation**
- Focused on fostering students' technical skills and in-depth knowledge in **statistical methodologies**, **experimental design**, **programming skills,** and applications to real-world problems

Akousist Technology, ***Data Engineer*** **Jun 2021 – Aug 2021**

Led analytical projects and enhanced database processes

- Collaborated with the data team to predict equity valuation using **Graph Convolutional Network** and text **sentiment analysis** using a **deep learning transformer**
- Developed **data pipeline** using **PySpark** and **Airflow** (DAG) integrating web scraping approaches from publicly available web pages leveraging word embedding vector models to ascertain keywords to achieve 83% prediction accuracy of pending lawsuits.
- Built dashboards using **Tableau** to analyze the prediction of trading based on sentiment analysis results in **Redshift**

AVerMedia Technology, ***Software Engineer*** **Jan 2019 – Jan 2021**

Managed large data volumes and processes and provide cross-functional support to both engineering and operations teams

- Collaborated with senior engineers to develop smart AI-supported retail and automotive systems to guide better business decision making that lead to higher revenues
- Improved **ETL** of large data sets (1 million+) using **PySpark (EC2 /EMR)** with **AWS S3** to speed up data processing and increase prediction reliability by 30%
- Implemented Model-View-Controller backend for third-party OAuth authentication with **Django** and construct **PostgreSQL** on the **AWS** to increase users' engagement and retention
- Utilized **shell scripting** in a **Linux** environment to device image processing tasks, such as image resizing, cropping, and format conversion, reducing manual effort and increasing efficiency

ARTIFICIAL INTELLIGENCE-BASED PROJECTS

Traffic Analytics – Vehicle Detection and Classification **Jun 2019**

- Created an image processing pipeline tailored to preprocessing large volumes of vehicle images with **OpenCV** and **FFmpeg** to enhance the performance of prediction
- Deployed **You Only Look Once v3 algorithm on TensorFlow** to detect, count and classify cars (with categories such as passenger cars, trucks, sport utility vehicles, etc.) on the road with to decrease traffic congestion by 10 %

Financial Analytics – **Currency Exchange Rate Forecasts** **Jan 2019**

- Developed an end-to-end data processing pipeline to scrape and clean currency exchange rate data from Federal Reserve System using **Beautiful Soup**, **Requests, Numpy** and **Pandas**
- Implemented **long short-term memory** algorithms to forecast the foreign exchange rate

EDUCATION

Georgia Institute of Technology **Expected: Aug 2023**

Master of Science - Computer Science (GPA: 4.0)

Relevant Course: Machine Learning for Trading, Database Systems Concepts and Design

University of San Francisco **Aug 2020 – May 2023**

Master of Science - Applied Economics (GPA: 3.8)

Relevant Courses: Data Structure & Algorithms, Natural Language Processing, Intro Statistical Modeling, Machine Learning, Experiments & Causal Inference