QIUSHI YAN

qiushi.yan@vanderbilt.edu ⬦ 629 239-9151

Website ⬦ Github ⬦ Linkedin

EDUCATION

Vanderbilt University — Nashville, TN
Master of Science in Data Science — 09/2021 - 05/2023

Communication University of China — Beijing, China
Bachelor of Arts in Data Journalism — 09/2017 - 06/2021

SKILLS

- 5+ years of experience in developing fullstack data-centric applications in Python, TypeScript and R
- Proficient in building Python and R libraries, focusing on extensible abstraction and architectural designs
- Skilled in database and DevOps on cloud such as DynamoDB and Docker, Certified AWS Developer ⬈
- An open source contributor familiar with git-based collaborations, CI/CD, and Agile workflows

EXPERIENCE

LearLab - Fullstack Engineer — 10/2022 - 11/2023

- Led the development of ITELL ⬈, a framework for building LLM-powered digital textbooks based on Next.js. Managed Node.js and Python SDKs.
- Developed model serving API endpoints in Python, with containerized deployment on Google Cloud.

RStudio - Software Engineer Intern — 05/2022 - 08/2022

- Developed the R package agua ⬈ for integrating the H2O.ai machine learning platform with tidymodels.
- Contributed bug fixes and features to other modelling packages in tidymodels: parsnip, tune, usemodels, bundle.

51Talk - Analytics Engineer Intern — September 2020 - June 2021

- Developed visualization dashboards for tracking model performance metrics.
- Rebuilt a searching service from legacy server deployment to serverless solutions with AWS Lambda.

DATA SCIENCE PROJECTS

agua ⬈ - Tidymodels Integration with H2O — R

- An R package interfacing tidymodels to the H2O machine learning platform, with a set of consistent APIs for model fitting, tuning, parallel processing, and interpretation.

Anime - A graph-based recommendation system — PySpark, PyTorch, Airflow

- A recommendation system for an anime streaming website, including PySpark ETL tasks in Airflow, training pipelines with Graph models in Neo4j, and microservice endpoints in FastAPI.

nyclodging ⬈ - An analytical dashboard for Airbnb pricings — R Shiny, JavaScript

- A machine learning web app to predict Airbnb pricing with word embeddings and multinomial regression.

SOFTWARE PROJECTS

xetra ⬈ - An ETL pipeline for trading data — SQL, Pandas

- Developed ETL pipelines to process the daily trading data of Deutsche Börse Group and save clean data to AWS S3. Fully tested with Pytest, deployed via AWS step functions.

Overwatcher ⬈ - An Esports API service — FastAPI, DBT

- Developed web scraping and data cleaning pipelines for game statistics in the Overwatch League. Served tidy data with FastAPI, Pandas, and PostgreSQL.

For more projects see projects.qiushiyan.dev.