QIUSHI YAN

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

Website ⋄ Github ⋄ Linkedin

EDUCATION

Vanderbilt University - Data Science Institute   Nashville, TN
Master of Science in Data Science   Expected May 2023

Communication University of China   Beijing, China
Bachelor of Arts in Data Journalism   September 2017 - June 2021

EXPERIENCE

LearLab - Fullstack Engineer   October 2022 - May 2023

- Developed the itell framework for building intelligent textbooks powered by LLMs.
- Built internal python packages for distributed data processing with Spark and Dask.

RStudio - Open Source Intern   May 2022 - August 2022

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

51Talk - Analytics Engineer Intern   September 2020 - June 2021

- Modeled key variables in teaching performance and ROI with regression and hypothesis testing.
- Developed a dashboard for monitoring key metrics identified above with Django and Vue.

DATA SCIENCE PROJECTS

agua - Tidymodels Integration with H2O   R Package Development

- Developed an R package interfacing tidymodels to the H2O platform and implemented a set of consistent APIs for model fitting, tuning, parallel processing, and interpretatation.

Anime Recommendation - A recommendation system for anime series   PySpark, PyTorch, Airflow

- Developed a recommendation system for an anime streaming website, including Airflow ETL tasks with PySpark, training pipelines with Node2Vec models in Neo4j, and forcasting services with prophet.

nyclodging - A predict web app for Airbnb pricings   Shiny

- Developed a shiny app generating geospatial visualizations and predicts price ranges for an Airbnb listing using word embeddings and multinomial regression.

SOFTWARE PROJECTS

xetra - An ETL pipeline for trading data   SQL, pandas, Pytest

- 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   data.table, FastAPI, DBT

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

For more projects see projects.qiushiyan.dev.

ADDITIONAL

Languages	R, Python, SQL, TypeScript
Frameworks	Shiny, pandas, PySpark, PyTorch, REST, GraphQL, DBT, React
Certification	AWS Certified Developer