Terry H. Ming

terru3@ucla.edu • 424-415-2534 • linkedin.com/in/terry-m3 • github.com/terru3

Education

University of California, Los Angeles (UCLA) Expected Graduation: June 2024
Bachelor of Science in Applied Mathematics; Statistics and Data Science Minor; Specialization in Computing
Cumulative GPA: 4.00

Skills

Programming Languages: Python *(pandas, NumPy, PyTorch, TensorFlow, Huggingface, scikit-learn)*, R, SQL, C++, HTML/CSS
Technical Skills: Git, Airflow, Snowflake, Streamlit, Tableau, Docker, NLP, A/B testing
Languages: English (Native fluency), Chinese (Native fluency)

Work Experience

Data Science Intern June 2023 — September 2023
Snowflake Inc.
Python, Snowflake, SQL, Git, Airflow, Streamlit, Docker

- Developed adaptive stratified survey sampling framework with multivariate testing, scheduled via Airflow, to iteratively calibrate and refine user experience models with human feedback
- Formulated mathematical framework for computing + updating account reputation scores via a combination of heuristics and behavioral anomaly detection
- Preprocessed and analyzed third-party data classifications, investigated use of crowdsourcing label aggregation algorithms to improve internal data labeling system

Pic 16B Reader: Python with Applications II January 2024 — March 2024
UCLA

- Read and graded 30 students' Python homework assignments

Leadership & Activity Experience

Data Science Union Project Member, Researcher October 2022 — Present

- Built and trained decoder transformer models in parallel from scratch on TinyStories, investigated scaling laws of dataset and model size with validation loss and created story generation demo
- Designed a transformer attention model leveraging patent long-form text to classify new patents into USPC categories and produce technology forecasts, achieving a top-5 accuracy of 81.6%
- Experimented with citation data and fine-tuning of Huggingface sentence transformers to improve patent classification performance

DataRes Researcher March — June 2022, January — March 2023

- Led a PageRank-centrality analysis graph project via Neo4j Graph Data Science and Cypher on 1,000,000 Spotify playlists
- Augmented a message-passing graph convolutional network with custom-defined socioeconomic indicators to improve traffic accident predictions

Association for Computing Machinery AI Member March 2022 — December 2022

- Developed a CNN in PyTorch to classify plant diseases
- Built a bidirectional LSTM with GloVe embeddings to identify insincere questions on Quora

Projects & Competitions

2022 DataFest Finalist + 2023
Python (pandas, NumPy, seaborn, plotly, scipy), R, Git

- Cleaned proprietary data, derived and presented insights in teams of five from challenging long-form datasets (100+ columns, 2+ million rows) in 40 hours
- Substantial use of seaborn and plotly visualizations, statistical tests, time series, survival analysis

LLMs for Question Answering

- Fine-tuned encoder-decoder model T5 with LoRA for extractive question answering on reading comprehension dataset SQuAD v1.1

Art Generation with GANs

- Implemented and compared DCGANs and Creative Adversarial Network (CAN)s to generate paintings, performed hyperparameter tuning and metric evaluations, developed interactive Streamlit demo

Rocket League E-sports Statistical Analysis

- Extracted 37,000+ series from public API, performed context-informed data wrangling and cleaning in pandas to obtain clear stat sheets for each player in every match
- Generated exploratory visualizations with seaborn, identifying interesting correlations to investigate
- Performed modeling/clustering to further analyze player behavior and uncover playstyle/team strategy insights

UCLA Hack on the Hill 9 (2022) [Education Category Winner]
Python (pandas, NumPy, beautifulsoup), Git

- Designed skeleton framework for a novel UCLA automatic degree planner with a team in 12 hours
- Wrote web-scraping algorithms incorporating regular expression matching to extract nested major requirements and prerequisite class data from various department and course catalogs