

KIEN TRAN

(470) 923-7528 | trantrikien239@gmail.com | github.com/trantrikien239 | master-data.science | Atlanta, GA

EDUCATION

Georgia Institute of Technology

Aug 2022 - Dec 2023

Master of Science in Computational Data Science

GPA: 4.0/4.0

- Coursework: Machine Learning, High Performance Computing (parallel algorithms), Information Retrieval Systems, Deep Learning, Large Language Models, Computer Vision, Operations Research, Regression Analysis
- Research assistantship: Developed ranking models on sparse healthcare data and restrictive computing resources

Foreign Trade University

Sep 2013 - Dec 2017

Bachelor's degrees – Double majors in Economics and Business Administration

GPA: 3.62/4.0

EXPERIENCE

Zillow Group | Applied Science Intern

May 2023 - Aug 2023

- **Developed customer preference machine learning models** using LightGBM and the Transformer architecture
- Utilized embeddings from pretrained graph and language models to enrich property and user representations, leading to a **3% overall and 8% key segment improvement** in roc_auc
- Provided **valuable insights for product enhancement** with model analyses (feature importance, ablation study)

One Mount Group | Lead Data Scientist (promoted from Data Scientist)

Aug 2019 - Jul 2022

- Spearheaded ideation, design, and development of **personalization and recommendation data products**
- **Led an agile data science team**, translating user stories to technical requirements, code reviews, mentoring team
- Extracted metadata from unstructured supermarket product description using BERT-based text classification and **named entity recognition (NER)** models, achieving **F1 scores of 0.86 for NER and 0.83 for classification**
- Developed and deployed a scalable lookalike engine using containerized Kubeflow components, enabling on-demand training and inference, resulting in **2.5x click-through rates** for marketing & advertising campaigns
- Designed and engineered a promotion recommendation system, leveraging uplift models, integer programming, and collaborative filtering to **boost transaction count by 90% and reduce excessive costs by \$35K**

Boston Consulting Group | Consulting Analyst

Jul 2018 - Jul 2019

- Performed data analysis for an Asian bank with **15 million customers**, resulting in actionable revenue strategies
- Identified clusters of customer personas through survey analysis and designed appropriate customer journeys

OpenCommerce Group | Data Analyst

Aug 2017 - Jun 2018

- Built and maintained an analytic database (HDFS, Presto) of 330K clients and **>10 million daily shopping customers**
- Prototyped a churn prediction model for a B2B e-commerce subscription product, reaching an F1 score of 0.85

PROJECTS

- **ChatGPT-powered Search Engine** ([GitHub](#)): Developed a search engine utilizing retrieval augmented generation (gpt-3.5 API, prompt engineering, Sentence Transformer) to create conversational responses with sources
- **Vietnamese Health Q&A Language Model** ([Report](#)): Procured a high-quality health Q&A dataset and employed LoRA and Quantization to fine-tune PhoGPT, reducing refusal rate by 100% and increasing helpfulness by 16%
- **Advanced Recommendation with Graph Mining** ([Report](#)): Implemented GraphSAGE & Graph Attention models using PyTorch Geometric for user-product link predictions, improved roc_auc by 2% over a strong NCF baseline
- **Generative AI for Inpainting** ([Report](#)): Implemented and evaluated 3 generative inpainting methods. Stable Diffusion was the most natural and outperformed DCGAN & Contextual Attention on FID

SKILLS

Programming: Python (Numpy, Pandas, scikit-learn, PyTorch), SQL, R, C++, Data structures & algorithms

Machine Learning: Random Forest, XGBoost; Linear & Logistics regression; Deep Learning: Vision (CNNs, ViT), Natural Language Processing (RNN, GRU, LSTM, Transformers), Graph Mining (Graph Convolution, Temporal Graph Networks).

Deployment: BigQuery, PySpark, MySQL, Google Cloud Platform (GCP), AWS, Kubeflow, Docker, Git, Shell scripting

Visualization: Pyplot & Seaborn; Microsoft Excel & PowerPoint; Google Data Studio, Tableau