Le Viet Thang

(+84) 989394637
Tan Binh District, Ho Chi Minh City
leviethang0512@outlook.com

Data Scientist / AI Engineer

GitHub: toreleon
LinkedIn: thanglv

I am a Data Scientist with nearly one year of working experience in production and over two years of working experience in Vietnamese NLP academic research. Experience working with NLP tasks such as dialogue system, information retrieval, machine reading comprehension, researching some fancy NLP topics, working on projects that are not in the NLP domain.

EDUCATION

Bachelor of Science in Data Science, *University of Information Technology*   Sep 2018 - Jan 2022

GPA: 8.21/10

Thesis: COVID-19 Vietnamese Machine Reading Comprehension with Numerical Reasoning

TECHNICAL EXPERIENCE

Data Scientist / Be Group JSC   **Jul 2021 — Present**

Customer Support Chatbot (PIC)

- Objective

  - Build a chatbot API service for the customer support team to collect information and reply to frequently asked questions from user and driver on automatically on the mobile application.
- Responsibility

  - Research in modern modularized and end-to-end dialogue system architectures.
  - Applied Rasa - an open-source conversational AI to create a chatbot system with historical dialogue from customer.
  - Wrote annotation guidelines and monitored the annotation process of the chatbot dataset.
  - Built and deployed an API service with FastAPI to the Kubernetes cluster.

Food Search Recommendation (PIC)

- Objective

  - Build a semantic search engine to recommend relevant restaurants and dishes based on the user's input query.
- Responsibility

  - Research in semantic search engine, spell correction and language model architectures.
  - Implemented a semantic search engine with a FastText language model using user search history data.
  - Designed an incremental training pipeline with Apache Airflow for automatically updating the semantic search model with new data daily.
  - Built and deployed an API service with FastAPI to the Kubernetes cluster.

Drop-Off Recommendation

- Objective

  - Build a recommendation system to recommend appropriate drop-off locations for users based on their current location.
- Responsibility

  - Research in building a personalize recommendation system with deep learning model.
  - Implemented a drop-off recommendation system using the ST-LSTM model with attention mechanism to rank the user's historical drop-off locations.
  - Supported API service development and deployment.

Research Student / DS@UIT Laboratory - University of Information Technology   **Feb 2020 — Feb 2022**

COVID-19 Vietnamese Machine Reading Comprehension with Numerical Reasoning

- Advisor
  - MSc. Nguyen Van Kiet
- Objective
  - Build a machine reading comprehension system integrated with numerical reasoning on COVID-19 news in Vietnamese.
- Responsibility
  - Research in machine reading comprehension model architectures and numerical reasoning in graph neural network.
  - Proposed the first numerical reasoning machine reading comprehension dataset, Vi-COVIDQA.
  - Wrote annotation guidelines and monitored the annotation process of the Vi-COVIDQA dataset.
  - Implemented SOTA numerical reasoning machine reading comprehension models such as NAQANet and NumNet to evaualate the performance on Vi-COVIDQA dataset.
  - Experimented and error analyzed to clarify the hypothesis in experiment results.

Research Assistant   **Oct 2021 — Dec 2021**

Real-Time Traffic Flow Prediction using Big Data Analytics (Accepted at ICIT 2022)

- Advisor
  - PhD. Do Trong Hop
- Objective
  - Build a forecasting system to predict the probability of traffic congestion based on the average velocity measured by multiple sensors on New York streets.
- Responsibility
  - Applied data science tools to pre-process large data in a limited memory environment.
  - Designed an incremental training pipeline for automatically updating the forecast model with new data daily.
  - Built a map to visualize the streamed traffic flow and the predicted flow in future time steps.

SKILLS

Tools and Languages	Python, Git, Markdown, C/C++, R, SQL, Docker
Data Science Toolkits	Numpy, Pandas, Tensorflow, Pytorch, Scikit-learn, Hugging Face, NLTK, Scipy, Matplotlib, Seaborn, Spacy, Gensim, PySpark, OR-Tools, etc
Quantitative Research	Mathematical optimization, Mathematical Modeling, Statistics and Inference
Workflow Orchestrators	Airflow, Pipelines
Data Warehouse Platform	BigQuery
Cloud Computing Service	Google Cloud Platform
Communication	Vietnamese (native), English (Intermediate)
OS Platform	Windows, Ubuntu