Nguyen Mai Chi Tan

HCM City, Vietnam • nmctan3056@gmail.com • 0931800483 • linkedin.com/in/nmctan6503
github.com/gamind65

Machine Learning Engineer with a passion for automating tedious tasks requiring much human resource with a proven track of developed an algorithm reduced manual labor by 45%. Specialized in Natural Language Processing, with experience in finetuning and inference LMs including LLMs. Eager to develop innovative solutions, explore new technologies, and learn about deployment processes.

EDUCATION

Bachelor of Computer Science **Aug 2021 - Jun 2025**
University of Information Technology, Ho Chi Minh city, Vietnam **GPA:** 7.87/10

PROJECTS

Aspect-based sentiment analysis for food reviews | Python, HuggingFace, Selenium, Streamlit
- Team size: 2
- Contribution:
  - Using Selenium to crawl up to 3500 food reviews across websites (i.e. foody.vn, tripadvisor.com.vn,...) then created fully labeled dataset with 1000 samples.
  - Recognized problem with auto-translated reviews, suggest and implement solution for auto-translated reviews by adding label for these samples thus improves models accuracy by 18%.
  - Preprocess data by textualize emojis, then using vncorenlp to tokenize sentences.
  - Finetune and evaluate PhoBERT and XLM-RoBerTa on created dataset with best accuracy at 48% then created a demo with Streamlit.

Abstractive News Summary - Automatic dataset constructing method | Python, HuggingFace, Pandas, Gradio
- Team size: 2
- Contribution:
  - Work with other team members to developed an automatic, scalable method to generate high quality Vietnamese news summary dataset, reducing human labor by 45%.
  - Designed a Bayes optimization function to calculate metrics' threshold for data filtering, eliminate the need of human manually written summary.
  - Implemented different metrics (i.e BERTScores, MINT,...) to filter data automatically while also ensure data high-quality.
  - Finetune, evaluate ViT5 model on our generated dataset, achieved 46.9% on ROUGE-L and 81.5% on BERT-Score Recall then created a demo with Gradio.

AWARDS

Top 8% in Home Credit - Credit Risk Model Stability by **Kaggle**
- Team size: 3
- Contribution:
  - Create a correlation matrix to visualize correlation in data and suggest which data feature should be removed.
  - Suggest and implement a weighted voting classifier (XGBoost, CatBoost, LightGBM) for making prediction, result in 29% increment in stability scores.

SKILLS

Programming Languages: Python, C/C++, SQL

Frameworks: HuggingFace, Pytorch, Tensorflow, FastAPI

Libraries: Scikit-learn, pandas, OpenCV, Matplotlib, seaborn, Selenium

Soft skills: Teamwork, read technical document, write technical report, scientific paper understanding.

CERTIFICATIONS

TOEIC Listening & Reading: 965
TOEIC Speaking & Writing: 320
DevOps, DataOps, MLOps by **Duke University - Coursera**
Applications of AI for Anomaly Detection by **NVIDIA**