Ho Chi Minh City, Viet Nam
github.com/tisu19021997

PHAM MINH QUANG

(+84) 769-009-285
quangphamm1902@gmail.com

EMPLOYMENT

Data Scientist **CADS - FPT Telecom** **Nov 2020 – Aug 2023**

FPT Customer Data Platform (CDP) Product

- Objective: Centralize customer data from various FPT subsidiaries to enhance customer relationship management and business intelligence.
- Customer Data Pipeline Development:
  - Engineered a production-grade, high-speed identity resolution algorithm in PySpark for over XX million customers across five diverse data sources.
  - Implement an evaluation score to ensure the output remain user-friendly and actionable for end-users.
- Product Design Leadership:
  - Spearheaded the initiation and development of prototypes for most product features.
  - Directly designed most UI/UX in Figma.
  - Guided the design team in creating user interfaces, data visualization, and customer data modeling

Long Chau Customer Insights

- Customer segmentation: Implemented customer segmentation using RFM analysis and K-Means algorithm, which was integrated to FPT CDP.
- Predict potential diseases: Utilized a combination of GPT-3.5 and open-source large language models, data crawling techniques, and BERTs to predict potential diseases in customers based on transactional data, achieving an 85% Jaccard Similarity compared to human annotation.

FPT Shop Customer Insights

- Established a daily data pipeline capable of processing XX thousands of transactions.
- PowerBI Dashboard: Develop an insightful dashboard to monitor store sales and customer return rate.
- LSTM Model: Trained an LSTM model with 90% accuracy to predict customer's gender based on their names.

Awards

- Most valuable player (MVP) of Q1/2021.
- Employee of the year 2021.

Software Engineer Intern **Kyanon Digital** **Sep 2018 – Mar 2019**

- Develop front-end for ecommerce website utilizing CSS, HTML, and PHP.

EDUCATION

Ho Chi Minh, Viet Nam **International University VNU** **2015 –2019**

- B.S.E. in Computer Science Engineering with Minor in Mathematics, 2019
- **Thesis Project:** Developed a comprehensive E-commerce platform, handling both front-end and back-end operations. The platform features an advanced Machine Learning-based recommendation system, enhancing user experience by providing personalized product suggestions.

TECHNICAL EXPERIENCE

Feedback Prize – English Language Learning (2022), kaggle.com/quangphm

- Objective: Evaluate essays by students on 6 analytic measures (from 1.0 to 5.0).
- Fine-tuning SOTA NLP transformer models like RoBERTA, DeBERTa, GPT-2. Further improve the score by ensembling 12 models and using different pooling heads (mean, attention, weighted, and mixture of them).
- Almost got silver medal but ended up choosing the wrong solution for the final submission (highest score 0.437311 but selected 0.435935).

Fine tune Llama-2 for better understanding Vietnamese, github.com/my-llama

- **Synthetic data generation:** Because of lacking high-quality Vietnamese instruction-following dataset and lacking resources to use OpenAI API, I built a Selenium agent to imitate user to insert prompt one by one and collect the response. Generated more than 5000 samples per day (but the duplication rate is high, about 40%). After about a week and with some post-processing steps, the result is high-quality 15K Vietnamese instruction-following dataset.

- **Fine-tune Llama-2: U**sed HuggingFace ecosystem (PEFT, Accelerate) and Kaggle free T4s to fine-tune 7B full-precision and 4-bit quantized 13B.

Full-stack Ecommerce web with recommender system, github.com/thesis-recsys

- Objective: Develop a complete Amazon-like Ecommerce website from scratch for undergraduate thesis and implement thesis' paper incremental SVD algorithm.
- Tech-stack: React (front-end), NodeJS (back-end), Flask (recommender system), MongoDB (database), Microsoft Azure (model storage).

PUBLICATION

- **Incremental SVD-based Collaborative Filtering Enhanced with Diversity for Personalized Recommendation (ICCCI 2020):** Matrix factorization-based recommender system using Incremental Singular Value Decomposition with Explicit Query Aspect Diversification.

Languages and Technologies

- Python; JavaScript.
- HuggingFace; Pytorch; Scikit-learn; Apache Airflow; NumPy; Pandas; Spark; Seaborn; Microsoft Power BI.