VY NGUYEN

Boston, MA | nguyen.vy7@northeastern.edu | Portfolio | LinkedIn | Github

EDUCATION

Northeastern University | Boston, MA   **Sep 2022 - Dec 2024**
Master in Data Analytics Engineering   GPA: 4.0/4.0
Coursework: Data Analytics, Data Mining, Neural Networks, MLOps, Statistical Learning, Data Management, Data Visualization

Foreign Trade University | Ho Chi Minh, Vietnam   **Sep 2013 - Sep 2017**
Bachelor of International Finance, CFA Level 2   *GPA: 3.3/4.0*

TECHNICAL SKILLS

Languages & Libraries: Python (Pandas, PySpark, PyTorch, Lightning, Numpy, Scikit-learn, Streamlit), MySQL, PostgreSQL, NoSQL
Machine Learning: Linear Regression, Logistic Regression, Random Forest, XGBoost, CNN, Transformer, Large Language Model
Technologies: Snowflake, Airflow, CI/CD, WandB, HuggingFace, Power BI, Tableau, Microsoft Azure, AWS, VS Code, Git, AnalyzeRe

WORK EXPERIENCE

Data Analyst Co-op | Vantage Risk | Boston, MA   **Jan 2024 - Now**
- Implemented a custom **data pipeline** in Azure to transfer data from our pricing tool to Snowflake. Developed **dbt models** based on logical models using the piped data and other existing databases as sources.
- Developed a **Streamlit** app connecting contract and loss data from Snowflake and AnalyzeRe API, to provide insights and analysis (e.g., NAV, seasonality analysis), including features like **geological maps**, and **parallel charts**
- Utilized Snowpark and SQL in **Snowflake** to gather and transform data, and created a real-time portfolio dashboard to display contract terms and modeled loss data, and perform risk analysis
- Created real-time reconciliation **Power BI** dashboard to assist the team in resolving data discrepancies for reporting

Graduate Teaching Assistant | Northeastern University | Boston, MA   **Sep 2023 - Dec 2023**
- Prepared course content, boosting engagement and comprehension for a 200-student class
- Held weekly office hours, resolving student queries and improving course outcomes

Data Scientist Intern | VNG Corporation | Remote   **Jun 2023 - Sep 2023**
- Enhanced current in-house auto **EDA** framework (Python) to further automate the data exploration process
- Developed a **chatbot** using **GPT-4 API** for household living standards inquiries, implementing **Prompt Engineering** to enhance performance by 30% over GPT-4 baseline

Financial Data Analyst | VietDragon Securities Corporation | Ho Chi Minh, Vietnam   **Jul 2018 - Dec 2019**
- Identified and communicated financial **trends and patterns** of 20+ companies to investors through **Tableau** dashboards
- Wrote **Python scripts** to automatically generate weekly stock price reports in HTML for broker team

Financial Analyst | Unilever | Ho Chi Minh, Vietnam   **Dec 2017 - May 2018**
- Developed strategic plans and conducted profitability, cost-benefit analyses to ensure alignment with corporate goals

ACADEMIC PROJECTS

Suicide Risk Detection on Social Media | IEEE BigData Cup 2024 [Github] [Competition]   **Summer 2024**
- Using **Large Language Model (LLMs)** for suicide detection: Prompting **Qwen2-72B-Instruct** to generate pseudo-labels for unlabeled data, and then fine-tune **Llama3-8B** and **Gemma2-9B** models with **LoRA**
- **Highest score** among 47 teams, based on *model performance* (6th on private board), *approach innovation*, and *report quality*

Breast Cancer Image Classification | Northeastern University [Github]   **Fall 2023**
- Re-implemented **ResNet** from scratch using **Pytorch** to classify histopathological images into malignant or benign cancer
- Delivered ablation study insights on architecture and training strategies, such as optimizers, number of layers, and batch norm

Term Deposit Prediction | Northeastern University [Github]   **Fall 2022**
- Analyzed a tabular dataset of 40,000 bank customers, addressing imbalanced data using oversampling methods (**SMOTE**)
- Achieved an F1 score of 0.89 with **LightGBM** for term deposit subscription prediction, improving from a baseline score of 0.65

Friends Episode Lookup | Personal Project [Streamlit]   **Spring 2023**
- Developed web-app using **Streamlit** and **ChatGPT API**, enabling to search for episodes of *Friends TV Show* via dialogue text

PUBLICATIONS

- **Vy Nguyen**, Chau Pham, "Leveraging Large Language Models for Suicide Detection on Social Media with Limited Labels", IEEE International Conference on Big Data 2024 [Paper]