Hong Van Pham

vanhongpham01@gmail.com • (438) 630-0695
malajvan.github.io/minivan • linkedin.com/in/vanhpham • github.com/malajvan

EDUCATION

McGill University **2019-Present**
Bachelor in Mathematics and Computer Science Montreal, Canada

- 3.57/4.0 GPA, Graduating May 2023
- Relevant Courses: Database Systems (COMP 421), Data Science (COMP 598), Applied Machine Learning (COMP 551), Artificial Intelligence (COMP 424)
- McHacks 10 (2023), MHCPP Volunteer, McGill Musicians Collective, La Belle Tonki (cook)

WORK EXPERIENCE

Software Developer **May 2022 – Present**
Distributed Digital Music Archives & Libraries Lab, McGill University Montreal, Canada

- Developed Rodan, a **Django**-based workflow management software for **Optical Music Recognition** and Analysis
- Built and deployed **Docker** containers on staging and production servers, reduced build time by 40%
- Identified and fixed infrastructural and functional user-reported issues and released new stable version 2.0.0
- Migrated codebase, **PostgreSQL** databases, and dependencies from **Python 2** to **Python 3.** Decreased average run time for different standard workflows by 15-30%. Reduced crash rate by 60% in 2 months
- Collaborated with different teams to ensure stability of new music analysis features releases and wrote detailed documentation

Data Science Intern **Jan - Apr 2021**
Vivas Technology Hanoi, Vietnam

- Eliminated 10% of office manual data entry tasks by implementing reusable **ETL pipeline** from factories to **ClickHouse** with **Apache Nifi** using **SQL**
- Built performance metrics BI dashboards in **Apache Superset** for the board of directors' decision-making
- Automated data loading tasks and cut data analysis time by an additional 30%

PROJECTS

Restaurant Review Analysis using NLP **Dec 2022**

- Gathered La Belle Tonki's 1000 Google reviews and applied NLP and **text processing** techniques including tokenization, punctuation and stop words removal.
- Applied **Random Forest** algorithm on customers' star ratings to classify reviews' sentiment and identify key subjects.
- Produced a report on customers' satisfaction and trends, visualized insights using **seaborn**

Spotify Popularity Predictions **Jan 2022**

- Extracted songs' audio features (MFCC) using **librosa** and **PyTorch** and predicted songs' popularity
- Implemented **KNN** algorithm with **pandas** and **scikit-learn** to classify popular songs' common features
- Correctly predicted 89% of sampled data in comparison to Spotify's API

Covid-19 Twitter Sentiment Analysis **Nov 2021**

- Queried 1000 tweets about COVID-19 in North America from twitter's API, tokenized and preprocessed texts.
- Developed 5 topics with a focus on vaccination. Calculated **tf-idf scoring** combined with **sentiment analysis** to characterize and produce a written report on general public's attitude during the pandemic

SKILLS & INTERESTS

- **Technologies & Skills:** Python, Azure, Docker, Git, Machine Learning, SQL, R, OCaml, Java, C, bash, Web APIs, Apache Superset, Apache Nifi, MatLab, Latex, PostgreSQL, Data Analysis, React, JavaScript, AWS
- Intermediate French, Fluent English and Vietnamese. Avid Musician and Vinyl Collector