**Tran Duc Trung**
Email: trung1803lucky@gmail.com
Phone: +(84) 325647395
Github | Linkedin

## 1. Goals and objectives

I thrive on coding and constantly seek out new challenges to explore; big data and AI are my interests. Over the next 3-5 years, my ambition is to launch my career as a data engineer while delving deeper into the realms of big data and AI. My goal is to master the intricacies of data engineering, all while expanding my knowledge and skills in AI and ML. With dedication and ongoing learning, I envision transitioning into roles like MLE or MLOps, where I can leverage my expertise to drive innovation and impact.

## 2. Education

**VNU-HCM University of Science**
*2021 – 2025 (Expected Graduation)*

- Bachelor of Data Science
- GPA: 3.4/4.0

## 3. Projects

**ETL Data Pipeline For Trip Record** Oct – Dec. 2023

- Motivation: Build an ELT data pipeline with the TLC Trip Record Data. Helping taxi businesses acquire sufficiently good and clean data before conducting statistical analysis or building models for prediction, thus deriving insights to enhance taxi services based on observed patterns within the data.
- Github: NYC-TripRecord with #Demo
- Tasks:
  - Ingest data into database.
  - Develop end-to-end data platform following Lambda architecture by build ETL pipeline, batch processing with Apache Spark.
  - Transfer data to data warehouse.
  - Make some analysts and demo data with a web app written by Streamlit.
- Technologies: Docker, Dagster, Apache Spark, MySQL, MinIO, PostgreSQL, Streamlit.

**House Price Prediction** May – June. 2023

- Github: House-Price-Prediction
- Tasks:
  - Using beautifulSoup and regex to crawl data on web
  - EDA, cleaning and preprocessing data (using MICE, one-hot encoding, scaling, remove outlier)
  - Build and compare to choose best model for predict house price (Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, Random Forest, CatBoost, XGBoost, Stacking model)
- Technologies: beautifulSoup, regex, scikit-learn, matplotlib, streamlit

## 4. Skills & Coursework

- **Programming Languages:** Python, C/C++, R, Matlab.
- **My tech stack:** Basic Python libraries (like Numpy, Pandas, Matplotlib, Seaborn, Streamlit ...), Scikit-learn, Pytorch, Spark, Dagster, dbt, MinIO, Docker, Linux (Ubuntu).
- **Database**: SQL (Microsoft SQL Server, PostgreSQL, MySQL), NoSQL (MongoDB).
- **Relevant School Coursework:** DSA, OOP, Probability and Statistics, Discrete Math, Databases, Database Management Systems, Intro to AI, Data Mining, Machine Learning, Pattern Recognition, NLP, BigData.
- **Languages:** English (intermediate).

## 5. Activities & Certifications

- Fundamental Data Engineering at #AIDE Institute
- VIASM The Summer School In Data Science 2023 at #VNU-HCMUS
- Google Cloud Skills Boost at #QuanQuanGCP
- HackerRank SQL (Basic to Advanced) Skills Certifications #HackerRank