To Duc Anh

toducanh2001@gmail.com
My LinkedIn: To Duc Anh

Hyprnx
</> Python, C/C++, Docker, R, SQL, HTML
See Dark theme

Introduction

I am a third-year Data Science student at National Economics University (NEU). With my childhood passion for numbers, economics, and programming, I believe that I am capable of finishing the assigned tasks at the finest level.

Experience

- **TNG Holdings Vietnam - TNG Loyalty Department** — Hanoi, Vietnam
  Collaborator Data Scientist — *April 2022 - Present*
  - Take part in analyzing data and building Data Lake, Data Warehouse.

  Technologies/Tools: Basic Knowledge of Networking, On Cloud and On-Premise, k8s, Ansible, Jenkins and Nexus. Designed and fabricated the above technologies on Linux (CentOS) Virtual Machine with Vagrant.

- **DSLAB NEU** — Hanoi, Vietnam
  Member — *November 2021 - Present*
  - DSLAB is a laboratory belonging to the Faculty of Mathematics Economics (MFE), National Economics University (NEU), specialized in the fields of Data Science and Artificial Intelligent (DS & AI).

- **IBM's Qiskit Localization Team** — (remote) Hanoi, Vietnam
  Translator/Proofreader — *March 2021 - Present*
  - Localize IBM's Qiskit documentation to Vietnamese for the community.

  Technologies/Tools: Python, Qiskit, Crowdin.
  Theory: Programming, Quantum Physics, Quantum Programming.

- **Vietsearch Foundation** — Hanoi, Vietnam
  Intern & Collaborator — *August 2020 - March 2022*
  - Designed, implemented crawlers to crawl data from LinkedIn, Google, Wikipedia, etc. with BeautifulSoup and requests library in Python.
  - Optimized existing crawlers, lower three times running time.
  - Developed query data APIs to serve customer's demand based on Elasticsearch search engine.
  - Designed unit and intergration tests for APIs.

  Technologies: Docker, Python, Swagger, Flask, Beautifulsoup, Selenium, Unittest, etc.
  Database: Elasticsearch, MongoDB.
  Theory: Crawling data, cleaning data, systematize data, Query, API, Testing.

Education

- **National Economics University** — Hanoi, Vietnam
  Major: Data Science in Economics and Business — *September 2019 – May 2023*
  - GPA: 3.53/4.0
  - Third prize in student scientific research contest
  - The second semester scholarship of the school year 2021
  - The first semester scholarship of the school year 2022
  - Achieved A+ (10.0) in Data Structure and Algorithm course
  - Achieved A+ grade in Machine Learning 1 course
  - Achieved A+ grade in Data Preparation and Visualization course

Courses

- **Data Stucture and Algorithm (DSA)** NEU, Hanoi, Vietnam

  OOP: Understand about four concepts of OOP
  Data Stuctures & Theories: Understand Time Complexity, Space Complexity, Recursion, and the concepts and usages of basic data structures such as Array, Queue, Stack, Tree, Heap.
  Sorting Algorithms: Experienced using Binary Search, Insertion sort, Heap sort, Bubble sort. Basic understanding of Radix sort.

- **Machine Learning 1 (Base on Stanford CS229)** DSLAB, NEU, Hanoi, Vietnam

  Maths: Understand how to construct the models and prove the algorithms mathematically.
  Models: Familiar with different types of regression models such as Linear, Ridge, Lasso, ElasticNet Regression, as well as classification models like SVM, K-mean, Decision Tree. Experienced Ensemble methods such as Decision Tree, Adaboost, CatBoost, xgBoost and LightGBM.

Following sections items are clickable on the pdf file.

Certifications

- **Certificate** Received Date

  (Name and score(s) are given, if possible below)

  - IELTS 7.0 *Issued April 2019*
  - The Internet and Computing Core Certification: 2520 *Issued December 2019*
  - BLOCKCHAIN MATHEMATICS AND COMPUTING *Issued July 2021*
  - Qiskit Localization Contributor *Issued May 2021*

Projects

- Used car prediction project: A personal project participated in a Kaggle Competition, to get the lowest Root Mean Squared Error(RMSE) when predicting the used car's price. NaNs in Numeric columns were replaced with mean. One-hot Encoding and Target Encoding were also used to encode categorical data. There were ten regression models used to train the data with the help of GridsearchCV. The best model was LGMBregressor, which archived the RMSE of 119k, rank fourth in the competition.
- Exploratory Data Analysis for Hotel dataset: Work with a group of three as the leader for the midterm project of the Data Preparation and Visualization course to create an EDA for the Hotel dataset. The group used Python libraries such as NumPy for calculation, Pandas for data manipulation, Matplotlib, and Plotly for plotting data and creating interactive charts. The final result was 9 on a 10 scale.
- Database Management System final project: Directed a team of four to design and implement a database to manage a telecommunication company using SQL as a final project for Database Management System (DBMS). The database allows the company to create plans to serve prepaid and postpaid customers, manage cash flow, and log errors. The design also allows users to create feedback and rating. The final result received a nine on a ten scale.
- Personal Website: I want to make an introduction website for myself, so I taught myself HTML and CSS to build one and then used Github Pages to publish my website.

Personal researches and publications

- **Used Cars Prices Prediction** Published on National Scientific Conference
  Hanoi, April, 2022

  ISBN: 978-604-358-602-2
  -Investigated and extracted used cars information that are selling on Vietnamese E-commerce site.
  -Constructed a Machine Learning model to help Vietnamese choosing the right used cars for their demands.
  -Conducted a dataset consists of all cars specifications (with variants) while doing Data Prepossessing.