

To Duc Anh

toducanh2001@gmail.com

My Phone: 0983691898

My LinkedIn: To Duc Anh

 Hyprnx

</> Python, PyTorch, Keras, Docker, SQL

See Dark theme

INTRODUCTION

Senior at National Economics University, majoring in Data Science. Detail-oriented, Mathematics, Artificial Intelligence, and Data Science enthusiast. Excellent at self-study, capable of finishing the assigned tasks with superb quality.

EXPERIENCES

• Citigo Software - AI team, Data Department

Hanoi, Vietnam

Intern Data Scientist

August 2022 - December 2022

- Take part in constructing text classification machine learning model to predict product's category utilizing product name.
- Design and implemented a package that be able to extract brand name from product name, base on fuzzy search, regex and Trie Data Structure. Published to Gitlab for internal use.
- Upgraded existing Address Suggestion module to improve suggestion accuracy based on existing module and Elasticsearch.

Technologies/Tools: Sklearn, Regex, Pytorch and Keras

• 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 integration 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: 3.58/4.0
- Third prize in student scientific research contest
- Three consecutive scholarship (Second semester of 2021, 2 semesters of 2022)
- Achieved A+ in Data Structure and Algorithm, Machine Learning and Data Preparation and Visualization course

COURSES

- **Data Structure and Algorithm (DSA)**

NEU, Hanoi, Vietnam

OOP: Understand about four concepts of OOP

Data Structures & 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 LGBMRegressor, which archived the RMSE of 119k, rank fourth in the competition.
- **Car Specification project:** An open-source project that was created with the primary purpose of saving time for the cars-related research community. The project offers the community the dataset that includes all car models and their variations that were mass produced from 1985 to early 2022. The project also comprised a Scrapy-based crawler that have clear instruction to re-crawl the data again if needed.
- **Text Classification project:** Directed a team of four to design and implement a classification module to classify product names into four distinct categories. The module was later deployed on Streamlit Cloud for demonstration. Models that are based on transformer architecture (sBERT and phoBERT) were used to calculate text embedding before passing to custom-made two layers Neural-Net for the classification task. The ONNX version of the models was used to accelerate the inference time.
- **Personal Website:** A website were made using HTML and TailwindCSS framework, with the main purpose of making a portfolio/introduction page for myself.

PERSONAL RESEARCHES AND PUBLICATIONS

- **Used Cars Prices Prediction**

Published on National Scientific Conference

Hanoi, April, 2022

Archived Third Prize in NEU Scientific Research Competition

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.