THONG HOANG VO

(+84) 522004060 | thongvhoang | thongvhoang | thongvhoang | vhthongvn | hoangthongvo

EDUCATION

- **VNUHCM - University of Information Technology** *Aug 2018 - Aug 2022 (Expected)*
  Bachelor of Science, major: Data Science, in-major GPA: 8.76/10.00 (3rd year)
  Coursework: Statistical Machine Learning, Deep Learning, Optimization and Applications, Data Analysis and Visualization, Experimental Design Analysis, Data Collection and Preprocessing, Data Structure and Algorithm, Object Oriented Programming, Probability and Statistics

- **Le Quy Don Khanh Hoa High School for the Gifted** *Aug 2015 - May 2018*
  High school education, class of Informatics

WORK EXPERIENCE

- **Officience**
  Machine Learning Engineer — Aug - Oct 2020
  - Crawled, optimized crawling process and pre-processed data from 5 real estate sites by **Scrapy, beautifulsoup4, lxml**
  - Performed **Bash script** that automatically inserts raw data into databases
  - Cleaned and wrangled raw data on a database using **Pandas**
  - Visualized data using **Matplotlib**, **Seaborn** and **PySpark**

RESEARCH EXPERIENCE

- **Information Systems Laboratory - University of Information Technology**
  Research Assistant — Jul 2020 - Current
  - Supervisor: Dr. Trong-Hop Do
  - Implemented object detection (**Yolov3**) and classification (**Xception**) algorithms for Vehicular Optical Camera Communication
  - Improved performance of object detection models on datasets
  - Techstack: Keras, Tensorflow

PUBLICATIONS

- Trong-Hop Do, Ngan-Linh Nguyen, **Hoang-Thong Vo**, Thanh-Binh Nguyen, and Khanh Ngo, "Deep learning based image processing for proactive data collecting system for autonomous vehicle", in 21st ACIS International Semi-Virtual Winter Conference.

PROJECTS

- **Predictive models for diabetes mellitus** *Feb 2021*

  Description:
  - A personal side project experimenting on multiple binary labels tabular data classification models.
  - Ranked in the top **47% percentile** of a data science competition on **Kaggle** with accuracy more than **85%** and the highest accuracy on that competition was over 87%

  Techstack: Scikit-learn, Numpy, Pandas, XGBoost

- **Classification of food for holidays and special occasions on the Epicurious - Recipes with Rating and Nutrition dataset** *Nov - Dec 2020*

  Description: Exploratory data analysis such as heat map and descriptive data statistics, handled missing data, experimented methods in feature scaling (Robust Scaler, Standard Scaler, ...), tried with classification models and deployed web app machine learning models

  Role: Team Leader

  Techstack: Scikit-learn, Matplotlib/Seaborn, Numpy, Pandas, Streamlit, Heroku

- **Analyzing the effect of these factors on students' final grades** *Jun - Jul 2020*

  Description: Exploratory data analysis, analysis of variance factors affecting final grades, regression analysis (Linear Regression, Support Vector Regression)

  Role: Team Leader

  Techstack: R, Python, Scikit-learn, Pandas

- **House numbers recognition on The Street View House Numbers Dataset using Convolutional Neural Networks** *Jun - Jul 2020*

  Description: Built the model which is able to recognize digits and numbers in natural scene images. The model was built on 2 dataset formats with 94% and 97% accuracy, respectively evaluated by the accuracy metrics.

  Role: Team Leader

  Techstack: Scikit-learn, Keras, OpenCV

- **Web Scraping** *Apr 2020*

  Description: Virtualized the Covid19 data from the Worldometer with InfluxDB. The model describes the total cases, daily new deaths for each country in the world.

  Techstack: Scrapy, Python, InfluxDB

- **Artificial Intelligence Nutritionist** *Sep - Dec 2017*

  Description: Diet Recommendation System for Healthcare Service Based on User Information using Neural Network in Deep Learning and FatSecret Food Database. This project awarded Encouragement Prize in the Science and Technology Contest for High School Students, Khanh Hoa Province, December 2017.

  Techstack: Python, Keras, Numpy, FatSecret Platform API

ACCOMPLISHMENTS AND ACTIVITIES

- **Honor prize in Provincial Informatics Olympiad for High School Students** *Nov 2016*
  Competed individually in 2 days to solve 3 data structures and algorithms problems in 3 hours on each day.

- **Gold medal in 30/4 Informatics Olympiad for Vietnam Southern High School Students** *Apr 2016*
  Competed individually in 1 day to solve 3 data structures and algorithms problems in 3 hours.

- **Member of Free Contest** *Feb 2016 - Feb 2021*
  These contests are for young coders in different regions in Vietnam to improve their coding skills with our high-quality problems.

LANGUAGES

- **Vietnamese**: Mothertongue
- **English**: Reading, Writing: Intermediate; Listening, Speaking: Basic

TECHNICAL SKILLS

- **Programming Language**: Python, C/C++, Bash script, SQL (Postgre), R
- **Machine Learning Algorithms**:
  - Supervised: Regression (Linear, Logistic), Memory Instance-based (k-Nearest Neighbor), SVM, Decision Tree (CART), Ensemble (Bagging, Random Forest, Boosting), Naive Bayes
  - Unsupervised: Clustering (K-Means), Linear Dimensionality Reduction (PCA, LDA)
- **Deep Learning Frameworks**: Scikit-learn, Keras, Tensorflow
- **Web Scraping Frameworks Proficient**: Scrapy, Request, Beautifulsoup, lxml
- **Data Analysis and Visualization Frameworks**: Pandas/Dask/Ray/Modin, Matplotlib/Seaborn, Scipy, Numpy
- **Version Control Proficiency**: Git, Github
- **Document Creation**: Microsoft Office Suite, LaTeX, Markdown

CERTIFICATIONS

- Computer Vision Basics - Coursera
- Fundamentals of Visualization with Tableau - Coursera