

Nguyen Cao Duy

🌐 ncduy0303.github.io | 📧 [ncduy0303](mailto:ncduy0303@gmail.com) | 📱 [caoduynguyen](https://www.linkedin.com/in/caoduynguyen) | ✉️ nguyen.cao.duy@u.nus.edu | ☎️ +65 9614 9992

EDUCATION

National University of Singapore

Bachelor of Science (Hons) in Data Science and Analytics

Aug 2022 - Present

Singapore

EXPERIENCE

A*STAR - Institute for Infocomm Research

Jan 2022 - Apr 2022

Student Research Intern

Singapore

- Crawled 1 million of English-Vietnamese sentences with BeautifulSoup and Selenium.
- Trained machine translation (40+ BLEU score - [Link to Personal Report](#)) and quality estimation models using Fairseq for Vietnamese Neural Machine Translation Engine.

PROJECTS

Object Detection and Audio Classification Models

Jun 2021

DSTA BrainHack TIL AI Camp

- Trained a YOLOv5 model to draw bounding boxes of animals in photos
- Trained an ensemble of ResNet, DenseNet, and VGGNet to classify animals' sounds by converting audio files into spectrogram photos
- Attained 1st Runner-Up overall in High School Category ([Digital Certificate](#))

Address Elements Extraction

Mar 2021

Shopee Code League - Data Science Challenge

- Trained most accurate named-entity recognition (NER) model (out of 1034 teams - [Link to Kaggle Competition](#)) using IndoBERT to extract Point of Interest (POI) Names and Street Names from noisy Indonesia addresses data provided by Shopee customers
- Attained 3rd overall after 3 challenges in Data Analytics, Data Science, and Algorithms ([Digital Certificate](#))

MathDoodle

Jan 2021

NTU SCSE Computing Challenge

- Developed most accurate Computer Vision model (out of 27 teams - [Link to Kaggle Competition](#)) to solve hand-written maths expressions
- Designed an interactive website using Streamlit to enhance children learning experience in mathematics ([Link to Github Repo](#))
- Attained top 5 overall and Young Innovator Award for creative approach

Other Projects

- **Product Detection** 🔗 : Trained an EfficientNet model to classify Shopee products from a noisy and imbalanced dataset (Kaggle Ranking: 23/824)
- **Email Campaign Analysis** 🔗 : Trained an XGBoost model to predict if users open emails sent by Shopee (Kaggle Ranking: 27/368)

AWARDS

- ASEAN Undergraduate Merit Scholarship
- National Olympiad in Informatics 2020, 2021 - Silver Awards
- Singapore Mathematical Olympiad 2020, 2021 - Silver Awards
- Singapore Physics League 2021 - Silver Award

TECHNICAL SKILLS

Programming languages: C++, Java, Python, R

ML/AI: Pytorch, Scikit-learn, Numpy, Pandas, Matplotlib

Web Technologies: HTML, CSS, Flask

Miscellaneous: OOP, CLI, SQL, Git, LaTeX

PROFESSIONAL DEVELOPMENT

AI: Deep Learning Specialization ([Digital Certificate](#)), Natural Language Processing Specilization ([Digital Certificate](#))

ORGANIZATIONS / CO-CURRICULAR ACTIVITIES

Problem Writer, Vietnamese Olympiad in Informatics Community:

2021 - 2022

Wrote 10+ competitive programming problems on VNOI Online Judge

Vice-President, Anderson Serangoon Junior College STEM Inc.:

2020 - 2021

Organised curriculum and workshop (e.g coding, flying drones, building robots) for members

Head Of Program Management, Bay Gio Tinh non-profit organisation:

2018 - 2019

Managed activities of an annual event by Vietnamese students overseas to teach local students soft skills