

Anh Vi Pham

Data Intern

Melbourne, Australia | anhvi09042002@gmail.com | +61 498 493 301

[Linkedin](#) | [Github](#) | [Kaggle](#)

SUMMARY

I am currently a penultimate year student at Deakin University, with nearly 4 years of exposure to the field of data science in both Vietnam and Australia. During this time, I have independently developed and honed most of my technical skills in data analysis through various personal projects. Despite my technical proficiency, I have identified a gap in my ability to analyze business problems effectively, an essential skill that, in my view, thrives in real-world scenarios. Consequently, I am seeking an opportunity in a professional environment where I can further develop my analytical capabilities and cultivate myself.

EDUCATION

Deakin University

Feb 2023 – present (expected graduation: May 2025)
Melbourne, Australia
Bachelor of Business Analytics

Da Nang University of Economics

Sep 2020 – Jan 2023
Da Nang, Vietnam
Bachelor of Data Science and Business Analytics
GPA: 3.55

SKILLS

- Data Manipulation: Python, SQL
- Machine Learning, Statistics
- Database Managing
- Cloud Computing: Azure
- Data Visualization: Tableau, PowerBI
- Web App: Django, Streamlit
- Data Scraping: Selenium, BeautifulSoup, Request
- Github, Microsoft Office

CERTIFICATIONS

- Hackerrank SQL (intermediate) certificate (Aug 2022) [view](#)
- Hackerrank Python (basic) certificate (Aug 2022) [view](#)
- IELTS 7.5 (2022)

PROJECTS

- Data GotTalent Contest 2024: Cinema Data Analysis and Dashboard** [dashboard](#) [github](#) Jan 2024
- Analyzed a dataset consisting of 40,000 records across 3 tables, revealing insights into business problems using statistical techniques in **Python**.
 - Designed and implemented a business database hosted on **Azure**, including an Extract, Transform, Load (**ETL**) pipeline.
 - Developed a **Streamlit** dashboard that displayed comprehensive information on 4,479 cinema members, along with total sales figures of 1,686,815,000 VND for 19 movies.
- Denmark Airbnb: Exploratory Data Analysis** [kaggle](#) [github](#) Nov 2023
- Analyzed data distribution, variable correlations, price differences, and customer satisfaction using **statistical** methods, revealing higher average satisfaction in Sealand compared to the rest of Denmark at a 95% confidence level.
 - Developed a classification model for apartment pricing using **scikit-learn**, achieving an accuracy of 75.3% and a recall of 74%.
 - Prepared an **analytical report** for management, summarizing key findings and insights.
- Amazon Product Analysis: Customers' Review Scraping and Sentiment Analysis** [github](#) Dec 2022
- Successfully scraped and analyzed 18,000 customer reviews from Amazon using **Beautifulsoup** and **Selenium**.
 - Preprocessed and vectorized text data using Natural Language Processing (**NLP**) techniques with **gensim**, **word2vec**.
 - Configured a Convolutional Neural Network (**CNN**) with **Tensorflow** to classify customer sentiments into 3 classes, utilizing a 5-layer architecture.
- Wine Quality Classification Model: Deployed with Django and Hosted on Azure** [github](#) Oct 2022
- Built and deployed a model for classifying wine quality using **scikit-learn** models, achieving a model accuracy of 83.25%.
 - Deployed the model using a web app created with **Django**, where users can input 9 features of wine bottles on Django to receive a quality prediction.
 - Hosted the web app utilizing **Azure App Service** for seamless accessibility.