

## Quoc-Anh Le

anh1q28102002@gmail.com | (+84) 978 207 750 | linkedin.com/in/davidle2810/ | davidle2810.github.io

### EDUCATION

---

#### HO CHI MINH UNIVERSITY OF SCIENCE - VNUHCM

##### *Bachelor of Science in Computer Science;*

Nov 2024

*Coursework:* Core Statistics, R Programming, Machine Learning, Artificial Intelligence, Natural Language processing, Mining on Big Data

#### HO CHI MINH UNIVERSITY OF SOCIAL SCIENCE AND HUMANITIES - VNUHCM

##### *Bachelor of Art in English Linguistics and Literature*

2026

### FEATURED PROJECTS (more details at <https://davidle2810.github.io/>)

---

#### Regression Techniques in House Price Prediction

- Performed comprehensive data analysis, data cleaning and feature extraction.
- Applied Stochastic Gradient Descent to build a regression model to predict house prices.

#### Recommendation System: Spotify Datasets

- Preprocessed a 100,000 entries datasets by handling missing values, duplicate values, unformatted text, unuseful columns.
- Analyzed the dataset and visualized interesting information.
- Applied PCA to reduce the numeric dimensions.
- Performed BFR clustering based on the mahalanobis distance.

#### Electrical Vehicle Charging Data

- Preprocessed dataset of 7,000 records.
- Conduct statistical analysis.

#### Frequent Itemsets with Association Rules Mining

- Performed product recommendation on the itemsets dataset filled with 100,000 entries using A-priori algorithm.
- Built the association rules for the item triples.

#### Sentence Alignment

- Predicted cross language by calculating the probability using softmax function with the application of BERT architecture and optimized the spans by using Integer Linear Programming.
- Built the model to align sentences in the Korean-English bi-text crawled from CNN and Yahoo.

#### Extracting Keywords using TF-IDF

- Performed keywords extraction in BBC news documents.
- Extract the top 5 keywords for each article and topic.

### SKILLS

---

**Programming:** Python (NumPy, Pandas, Scikit-learn, TensorFlow, nltk), R, SQL

**Visualization and Statistical Software:** Tableau, Python (Matplotlib, Seaborn)

**Machine Learning:** Regressions, SVM, XGBoost, Unsupervised Learning (Clustering, PCA), Deep Learning