QUAN VAN NGUYEN

⌂ Ho Chi Minh City
✉ harryquan.minh@gmail.com
○ GitHub Profile: minhquan6203

☎ (+84) 969 399 209
✉ 21521333@gm.uit.edu.vn
in Linkedin Profile: Harry Nguyen

EDUCATION

• **University of Information and Technology - Vietnam National University HCM**
Bachelor in Data Science
GPA: 3.2/4
Interest: Data Engineering, Cloud Computing, DevOps, MLOps, AI system.

SKILLS

Programming: Python(proficient), SQL(prior experience), C/C++, C#, R.
Frameworks: Spark, Kafka, Airflow, Pytorch, Tensoflow.
Tools: Azure, VSCode, Jupyter Notebook, Terminal, Git, Docker, PowerBI.
Knowledge & Course: Data Structure and Algorithm, Big Data, Cloud Computing, Distributed Database, Basic Software, Machine Learning & Deep Learning, Calculus, Linear Algebra, Discrete Math, Probability & Statistic.
Soft skills: English, management skill, teamwork, have experience in research with serveral publications.

PROJECTS

•**Real-time Sentiment Analysis for Fast Food Brand**   05/2023 - 06/2024
Role: Leader - Team size: 4
  – Tool & Technology: Kafka, Spark, Docker, Python, MongoDB, Natural Language Processing.
  – Developed a real-time sentiment analysis system to monitor customer feedback for a fast food brand.
  – Utilized Kafka for real-time data streaming.
  – Employed Apache Spark for data processing and analysis.
  – Deployed the solution in Docker containers for scalability and portability.
  – Ensured continuous data flow and real-time insights into customer sentiment, enabling quick response to feedback
  – Source code **HERE**

•**MLOps NY Taxi Fare Prediction on Azure**   04/2024 - 05/2024
Role: Leader - Team size: 2
  – Tool & Technology: Azure, Azure ML, Azure DevOps, Azure Storage, Python, Docker.
  – Implemented ETL processes and data preprocessing.
  – Created model training and deployment pipelines.
  – Utilized Azure DevOps for CI/CD.
  – Deployed the model to a Kubernetes (K8s) cluster for production.
  – Automated end-to-end workflows, ensuring scalable and reproducible ML operations.
  – Source code **HERE**

•**Airflow ETL Project**   04/2024 - 05/2024
Coursera project
  – Tool & Technology: Apache Airflow, Data Analysis, Data Preprocessing, Docker, Python.
  – Designed and implemented ETL pipelines using Apache Airflow.
  – Automated data extraction, transformation, and loading processes.
  – Scheduled and monitored workflows for timely data processing.
  – Source code **HERE**

ACHIEVEMENTS & CERTIFICATE

•**Certificate of ETL and Data Pipelines with Shell, Airflow and Kafka [HERE]**
Course completed with IBM on Coursera, Grade: 93%.   *06/2024*
•**Second prize - AI Challenge: Facial Analysis [HERE]**
Organized by Pixta Vietnam, Co-Host AI and Viblo.   *02/2024*
•**Third prize - VLSP2023: Legal Textual Entailment Recognition [HERE]**
Organized by VLSP - Vietnamese Language and Speech Processing   *12/2023*
•**First prize - Group B DSC 2023: Vietnamese Fact Checking [HERE]**
Organized by Faculty of Information Science and Engineering, University of Information Technology   *11/2023*

PUBLICATIONS

•**ViTextVQA dataset [HERE]**
Preprint   *04/2024*
•**ViOCRVQA dataset [HERE]**
Preprint   *04/2024*
•**MAT: Effective Link Prediction via Mutual Attention Transformer [HERE]**
Accepted by Data Science Advanced Analysis Conference (DSAA)   *11/2023*

LANGUAGES

English
Toeic
Reading - Listening: 620

<div align="right">08/2023</div>

PREFERENCES

•**MSc. Kiet Van Nguyen**
Vice Dean Faculty of Information Science and Engineering
University of Information and Technology
kietnv@uit.edu.vn

•**PhD. Trong-Hop Do**
Faculty of Information Science and Engineering
University of Information and Technology
hopdt@uit.edu.vn