

Khoi Mai

Santa Ana, CA | github.com/khoimai05 | (657)-565-2480 | maixuankhoi05@gmail.com

Education

Purdue University

Bachelor of Science in Computer Science and Data Science

May 2027

GPA: 3.63

Technical Skills

Libraries: Scikit-learn, TensorFlow, PyTorch, Pyspark, Pandas, Numpy, Seaborn, Matplotlib.

Technologies: Git, Docker, AWS, Airflow, SQL, Power BI

Experience

Discovery Park

West Lafayette, IN

Data Science Researcher

Dec 2025 - Now

- Performed data cleaning, joining, and engineered features from **200K+** EV charger permits (Shovels), **1M+** of POI visit records (Dewey), and U.S. Census data, enabling ad-hoc analysis.
- Conducted statistical analysis of **150+** CA neighborhoods linking home-charger penetration to travel/fueling behavior patterns.

Antsomi CDP

Ho Chi Minh City, Vietnam

Data Scientist Intern

Jun 2025 - Aug 2025

- Built a recommendation system for Mega Wholesale, serving **10K+** users with personalized suggestions and cross-selling insights, analyzing **\$1.2M** in annual revenue.
- Applied BERT embeddings and k-means clustering to categorize product groups, and performed association rule mining with Apriori across **30K+** transactions.
- Improved customer acquisition model processing **3M+** monthly interactions from S3 by engineering engagement interval and usage recency features.
- Achieved **89% accuracy** and **77% recall**, identifying **25K+** high-value prospects monthly from **300K+** users for Golden Gate Restaurant Group.

VRAI Lab

West Lafayette, IN

NLP Research Assistant

Sep 2024 - Now

- Engineered automated Python scripts ingesting **10,000+** scientific papers (20GB+ per run) from PubMed, processing **450+ papers/hour** with **86%** extraction success rate.
- Delivered text analysis service to Purdue faculty by applying LDA topic modeling and hypothesis testing to validate research themes.

Orient Commercial Joint Stock Bank

Ho Chi Minh City, Vietnam

Analytics Engineer Intern

May 2024 - Aug 2024

- Cleaned, transformed, and normalized **35,000+** records across multiple data sources; reduced processing time significantly to 3 hours.
- Optimized database architecture to **3NF** through Python scripts, automating schema normalization, achieving **30%** faster SQL query execution.
- Developed Power BI dashboards, visualizing **5+** market KPIs across **30+** provinces and cities to support market analysis.

Awards and Projects

Microsoft GenAI Hackathon Vietnam

Runner Up

June 2024

- Won **2nd** place out of 10 participating leading financial services companies.
- Led a team of **4** to create a virtual assistant to promote and market **12** credit card products.

Elite Edge

Basketball Analytics Platform

Dec 2025 - Now

- Constructed ML models (KNN, XGBoost) analyzing 12 years of data with **30+** features to predict tournament team selection and Final Four advancement.