Truong Tien Anh

+(84) 982 619 731 | truongtienanh16@gmail.com | linkedin.com/trgtanhh | github.com/trgtanhh04

EDUCATION

University of Science - VNUHCM   Oct. 2022– Present
Bachelor of Science in Data Science and Computer Science   *Current GPA: 3.7/4.0*

CAREER OBJECTIVE

I am a third year Computer Science student with a passion for data science, machine learning, and big data. I am looking for an internship or job opportunity as a Data Engineer or AI Engineer to apply my skills to real-world problems and grow through hands-on experience.

SKILLS

Programming Languages: Python, SQL, JavaScript, C/C++
Data Science & ML: LangChain, TensorFlow, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
Data Engineering: Apache Hadoop, Spark, Airflow, Kafka, PostgreSQL
Cloud & DevOps: AWS (S3, Glue, Redshift, Athena), Docker, Git/GitHub, MongoDB Atlas

CERTIFICATION & BLOG

TOEIC: 810/990
My Portfolio: portfolio.com/trgtanhh

PROJECTS

End-to-End MovieDB   February 2025 – Present
- **Key Technologies**: Apache Airflow, Spark, Kafka, HDFS, PostgreSQL, Streamlit, Machine learning
- **Description:**
  * Developed a complete data pipeline to crawl, process, and analyze movie data for powering a recommendation web app.
  * Crawled movie information and transformed it into structured JSON format.
  * Stored raw data in HDFS; used Spark for data cleaning and preliminary analysis.
  * Integrated Kafka and Airflow to orchestrate and trigger ETL workflows.
  * Loaded processed data into PostgreSQL, deployed on Neon for cloud-based access.
  * Built a prediction model for movie pricing and a Streamlit app for personalized recommendations.
- **GitHub:** github.com/trgtanhh04/End-to-End-MovieDB-Data-Engineering

Mobile AWS Pipeline Engineering   March 2025 – Present
- **Key Technologies**: AWS (S3, Glue, Athena, Redshift), Docker, Apache Kafka, Spark, PostgreSQL, Machine learning
- **Description:**
  * Built an AWS-based data pipeline to support a mobile recommendation system and price prediction model.
  * Crawled and ingested mobile phone data using Kafka, with event logs stored in PostgreSQL.
  * Processed streaming data in real time with Spark and stored outputs on S3.
  * Automated data cleaning via AWS Glue; queried processed data using Athena and Redshift.
  * Improved data processing speed by 30–50% through optimized infrastructure and parallel processing.
  * Designed dashboards to visualize key insights for decision-making.
- **GitHub:** github.com/trgtanhh04/Mobile-AWS-Pipeline-Engineering

CV Analysis using LangChain   May 2025 – Present
- **Key Technologies**: FastAPI, LangChain, OpenAI API, Streamlit, PostgreSQL, FAISS
- **Description:**
  * Built a system to extract and embed structured data from CVs using LLMs.
  * Stored data in PostgreSQL and FAISS for semantic search.
  * Developed search APIs to match candidates by job titles and skills.
  * Deployed backend (FastAPI) and frontend (Streamlit) to the cloud.
- **GitHub:** https://github.com/trgtanhh04/CV-Analysis-using-Langchain.git