Dai Vo - Data Engineer

☐ +84 944 433 047 | ⚏ 2000 | @ daivt1805@gmail.com | 🔗 LinkedIn | ⓞ GitHub | 📍 HoChiMinh, VietNam

OBJECTIVE

Detail-oriented data engineering with a strong focus on cost optimization, seeking a challenging role to design and implement efficient data pipeline. Committed to leveraging cloud-based technologies, automation, and data modeling expertise to deliver cost-effective solutions and optimize resource utilization. Then, analyze, and interpret data to provide valuable insights and support informed decision-making within an organization through data storytelling.

EDUCATION

University of Information Technology, VNU-HCM   HCM, VietNam
*B.Sc. in Data Science; **GPA: 8.51/10 Good***   *Sep 2019 – Jun 2023*

SKILLS

Programming: Python, C, C++, MySQL, Git.

Data Engineering: Spark, Kafka, Hadoop, Airflow, Docker, AWS.

BI tools: Grafana.

Soft skill: Life-long learning, critical thinking, logical reasoning, research, problem-solving.

Languages: Vietnamese (Native), English (Limited working).

AWARDS & ACHIEVEMENTS

Semester Scholarship: Ranked 3rd in the class with the highest overall score in the 1st semester.

CERTIFICATES

Machine Learning Specialization   *Feb 2022*
IBM Data Science Specialization   *Aug 2022*

PROJECTS

Covid-19 End-to-End Data Analysis
Build End-to-End data pipeline from data source to dashboard to display total cases, deaths and total cases per region.
- Automate pushing data from data sources into data lake.
- Build ETL pipeline to handle data.
- Build a data warehouse to save processed data.
- Build a real-time dashboard to display some insight of data.
- **Skills:** Python, Airflow, AWS (S3, Crawlers, Glue, Athena, Redshift, QuickSight).

Vehicle Speed Estimation on the Vietnam's Street Real-time
Video from data sources are pushed into model by Kafka. The vehicles are detected and tracked to estimate speed real-time.
- Train models on multiple GPU.
- Use Kafka to send the frames of video from data source into model.
- Write speed estimation of the vehicles based on the distance of the crossing-lines.
- **Skills:** Python, Kafka, OpenCV, Flask, Models (Yolov7, DeepSORT).

Hate speech detection on Vietnamese Social Media Text on Facebook Real-time
Comments from select public Facebook pages are crawled and pushed into Spark Structured Streaming to process them. Deep learning models trained using distributed methods will predict sentiment of comments. A real-time web app dashboard will display aggregated predictions.
- Use BigDL to improve performance by combining the power of distributed clusters when training models.
- Crawl data from Facebook by Selenium.
- Use Kafka to push data into Spark.
- Handle data real-time with Spark Structured Streaming.
- Build a web-app dashboard to display statistics of data and visualize prediction.

- **Skills:** Python, Selenium, BigDL, Keras, Spark, Flask, Plotly, Models (LSTM, GRU, Text-CNN, NaiveBayes, Logistic Regression, Decision Tree).

Real Estate analysis

Analysis price of real estate in Ho Chi Minh, Binh Duong, and Dong Nai from the posts on Chotot.com, Alonhadat.com.vn.

- Crawl data from Chotot and Alonhadat by Scrapy.
- Preprocess data with handling missing, outlier values.
- Execute Exploratory Data Analysis and Feature Extraction to take insights from data.
- Build machine learning models to check the hypothesis.
- Create web-app to visualize prediction from inputs
- **Skills:** Python, Scrapy, BeautifulSoup, Flask, Models (Linear/Polynomial Regression, Random Forest, KNN).