Trang Le

tlee.bon@gmail.com • ✆ 651- 460-0817 • linkedin.com/TrangLe • github.com • Portfolio

Skills

Languages: Python, SQL, Java, R

Databases: MySQL, MSSQL, MongoDB, PostgreSQL.

Tools/ Libraries/ Technologies: Git, Github, Gitlab, Docker, FastAPI, REST APIs, OpenCV, Pandas, Numpy, Keras, Selenium, Microsoft Office, RStudio, Airflow, Cloud Composer, AWS, GCP, Big Query, Jams, Kafka, PySpark, GraphQL.

Education

2023 **M.S. in Data Science**, University of Cumberlands, KY

2017 - 2020 **M.S. in Software Engineering**, University of St. Thomas, MN

Experience

08/22- Present **Data Engineer - Jewelers Mutual Group, TX**

- Building and maintaining scalable ETL data pipelines and data products using Airflow.
- Creating automated workflows using Jams scheduler that are ingested into GCS buckets as part of ETL processes.
- Collaborating with cross-functional stakeholders and engineers to design, develop, test, implement, and support technical solutions.
- Supporting and troubleshooting production issues on the data lake side.

09/21-07/22 **Data Engineer I - Flexshopper, FL**

- Developed a Python-based production micro-service using web framework FastAPI.
- Collaborated with the team to build our next generation of business intelligence suite and perform ad hoc data analysis using SQL and MongoDB to analyze high-volume, high-dimensionality data from various sources.

01 - 09/21 **Software engineer intern - Smartcare Software, WI**

- Integrated a background checks API into the platform and managed the process from end-to-end. Performed testing as well as monitoring in production environment.
- Assisted with migration of legacy systems and helped the team design, develop, and communicate modern infrastructure and processes. Worked on the backend with authentication and authorization.

Projects (Portfolio)

06/2021 **Random names streaming/ Python**

- This end-to-end project is streaming data (json formatted) from the Random Names API to Kafka and writing data to Bigquery table. This workflow is containerized within Docker and Airflow is used for scheduling the streaming jobs.

01/2020 **Cancer Predictor App/ Python**

- Performed an end-to-end lifecycle of Machine Learning development project. Achieved an impressive 90% accuracy.
- Performed a variety of **data analysis** techniques. Built models with: **SVM, Logistic Regression, Naive Bayes, KNN, XGBoost and Decision Tree**. Employed **GridSearchCV** for **hyper-parameters optimization** and **KFold cross validation** for models evaluation.

Certificates

Deep Learning Specialization, a 5-course specialization, Coursera

- Learned foundations of Deep Learning and Neural Networks. Worked on case studies from healthcare, autonomous driving, sign language reading, music generation, and natural language processing.

Publications

- **Le, T.Q.T**. and Rege, M. "Effectiveness feature for micro-expression recognition." Proceedings of the 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI). IEEE, 2021. **(oral presentation) .** [link]
- **Le, T.Q.T**., Tran, T.K., and Rege, M. "Dynamic image for micro-expression on region-based framework." Proceedings of the 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI). IEEE, 2020. **(oral presentation) .** [link]