LONG NGUYEN

ngkhlong189@gmail.com • 309-631-4516 • Linkedin • Portfolio

EDUCATION

UNIVERSITY OF ROCHESTER   Rochester, NY
STEM Master of Science in Data Science   Aug 2022 - Dec 2023
- Relevant courses: Data Mining, Statistics, Machine Learning, Database Management, Deep Learning.

AUGUSTANA COLLEGE   Rock Island, IL
Bachelor of Arts in Applied Mathematics   Aug 2018 - May 2022

PROJECTS

Data Analytics: Behavioral Segmentation for a hotel and resort company
- Built a K-means clustering model in **Python** on a hotel and resort dataset of > 80k customers to segment them into different behavioral groups, and provided key strategies to enhance customer experience and maximize revenue for each segment.

Machine learning: Mental health classification (Python) | Alternative implementation in R with statistical analysis
- Deployed 5 predictive models in Python to determine whether a tech employee sought mental health treatment, with SVM attaining a leading F1 score of 75% and an accuracy of 74.6%.

Data Engineering: NYC bike tracking forecast
- Led a team of four in developing an end-to-end ETL pipeline, employing PySpark to create bronze, silver, and gold data tables; Achieved a 70% accuracy rate in building a forecasting model using Prophet for the Citibike Bike Sharing app on Databricks.

Data visualization: Visualization of layoffs across industries
- Created an interactive dashboard with descriptive statistics and a world map showcasing layoffs across industries in Tableau.

EXPERIENCE

REL8ED.TO ANALYTICS   Rochester, NY
Data Science Intern   Jul 2023 - Aug 2023
- Led a project using Python and OpenAI's embedding model to classify >100K N.A companies into their respective industries with 85% accuracy, saving over $500K in manual process and laying the groundwork for future LLM implementations.
- Developed a multithreaded scraper for 10M companies' homepage data, stored in JSON format, collaborating with the software team to create a seamless API, resulting in a substantial 90% reduction in data retrieval time.
- Collaborated with the data team to deliver 6 comprehensive sales performance analysis reports in Excel tailored exclusively for clients, resulting in actionable insights for driving revenue growth, with a client satisfaction rate of 9/10.

LEAFPRINT ENTERPRISES   Cedar Hills, UT
Data Analyst Intern   Jan 2022 - May 2022
- Led a data research project on global municipal-level districts, overseeing data collection, resulting in a 20% increase in latency through Python automation scripts.
- Developed intricate dictionaries and efficient data models using SQL for World Archive's geographical data, contributing to a 30% enhancement in data organization efficiency.
- Created detailed documentation for Python scripts and SQL queries, facilitating smoother collaboration and enabling seamless project understanding, ultimately leading to a 40% reduction in onboarding time for new team members.

BENIT PTY LTD   Hanoi, Vietnam
Data Analyst Intern   Mar 2021 - Jul 2021
- Co-led a $5k Algo Trading project with a team of 4 Data Science interns; created an automated Python script to transform Forex's crypto data, and utilized SSIS to design and build an ETL pipeline for data capture.
- Facilitated seamless communication and cross-departmental alignment between 3 teams (data, software, business) to create an optimized trading portfolio with data insights.
- Successfully built multiple dynamic Tableau dashboards featuring key metrics like Average Daily Trading Volume and ROI, resulting in a 25% processing accuracy boost and 30% faster processing.

TECHNICAL SKILLS

Tools: Programming Language: Python (numpy, pandas, matplotlib, scikit-learn), SQL, R (tidyverse, dplyr, ggplot2, caret); Data Visualization: Tableau, Power BI, Excel; Cloud: AWS (S3, Redshift); Big Data Technology: Spark, Databricks
Skills: Statistical Analysis, Predictive Modeling, Regression Analysis, Machine Learning (unsupervised: Clustering; supervised: Classification, Prediction), Data Modeling