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, Comp. Statistics, Statistical Machine Learning, Database Management, Deep Learning.

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

EXPERIENCE

REL8ED.TO ANALYTICS   Rochester, NY
Data Science Intern   Feb 2023 - Aug 2023
- Led an end-to-end LLM project to classify North American companies into their respective industries, employing **Python**, OpenAI embedding model, and utilizing the ChatGPT chatbox function with optimized prompts.
- Achieved 90% in data retrieval time by building 5 **Python** scrapers to extract homepage information for 10M companies.
- Collaborated with business teams to convert complex business logics, and successfully built 6 **Excel** dashboards incorporating pivot tables, XLOOKUP based on client requests, reducing delivery time by 70%.

UNIVERSITY OF ROCHESTER   Rochester, NY
Teaching Assistant   Aug 2023 - Dec 2023
- Mentored students in statistical areas: descriptive statistics, probability, hypothesis testing, ANOVA, and regression analysis.
- Led multiple review sessions for exams and projects, resulting in an 80% increase in students' overall grades.
- Empowered students in understanding **R** packages and creating impactful reports with **RMarkdown**.

LEAFPRINT ENTERPRISES   Cedar Hills, UT
Data Analyst Intern   Jan 2022 - May 2022
- Led global municipal research, collected 100k+ data points, and delivered actionable recommendations to enhance World Archives's municipal dictionary.
- Utilized **Python**'s pandas for data cleaning and transformation, reducing errors and retrieval time by 90%.
- Developed and maintained interactive dashboards and reports using **Tableau** to support data driven decision-making.

BENIT PTY LTD   Hanoi, Vietnam
Data Analyst Intern   Mar 2021 - Jul 2021
- Developed an optimized **SQL** data model for TwelveAPI's Forex data, improving access time by 80%.
- Utilized **AWS S3** bucket and **MongoDB** for scalable database management, resulting in a 70% reduction in overall latency.
- Integrated **Python** with **SQL** databases to streamline **ETL** processes, ensuring accurate financial insights via dynamic dashboards and reports.

PROJECTS

Machine learning: Mental health classification
- Built 5 predictive models in **Python** to predict whether a tech employee sought mental health treatment or not, with SVM attaining a leading F1 score of 75% and accuracy of 74.6%.

Data Engineer: NYC bike tracking forecast
- Achieved 70% accuracy in building a forecasting model using **Prophet** for Citibike Bike Sharing app on **Databricks**.
- Collaborated on **GitHub**, developed end-to-end **ETL** pipelines, and created bronze, silver, and gold data tables using **PySpark**.

Data visualization: Asian hate sentimental analysis
- Utilized **Python** snscrape to scrape 16,000+ Asian and COVID-related tweets, conducting EDA and employing WordClouds for visualizing the most frequent hashtags.

TECHNICAL SKILLS

- **Programming:** Python, R, SQL
- **Big data processing:** Databricks, PySpark
- **BI technologies:** Tableau, PowerBI, Excel
- **Database:** MySQL, PostgreSQL
- **Proficiency**: queries & joins, data modeling, statistical analysis, predictive ML modelings, Jira (sprint)