Haoxin Luo

918-633-7217 | luohaoxin943@gmail.com | 1028 W 87th PL Tulsa OK 74132 | https://www.linkedin.com/in/cocoluo1999/

EDUCATION

University Of Washington — **Seattle, WA**
MS in Advanced Method and Statistical Analysis — *09/2021 - 06/2023*

Wake forest University — **Winston-Salem, NC**
BS in Statistics and Probability, BA in Computer Science — *08/2017 - 05/2021*

SKILLS

Fullstack Devlopement: Javascript, Typescript, HTML, CSS, React, node.js, express.js, PostgreSQL, Git
Analytical skills: Python, MySQL, R, Power BI, Tableau
Machine Learning: Regressions, SVM, Random Forest, Boosting, KNN, PCA, K means, Neural Network, Bayesian (MCMC)
Data Warehouse / Cloud Technologies: Snowflake, Azure, AWS

PROFESSIONAL EXPERIENCE

Occlusion — **Tulsa, OK**
Machine Learning Engineer — *09/2023 - present*

- Conducted **ETL** using **Python** on orphan wells data from RRC and US Department of Interior and built machine learning models for well ranking and prioritization to maximize residual value
- Imported cleaned data into Dbeaver and crafted a latitude/longitude map integration using **Snowflake SQL**
- Tailored the model for process optimization, cost management and regulatory needs
- Developed a well proximity planner website with **React** for the frontend and **Node.js** for the backend
- Collaborated with project manager to streamline well management, resulting in a 15% boost in productivity and a 20% improvement in financial efficiency

Eli Lilly and Company — **Indianapolis, IN**
Data Analyst Intern — *06/2022 - 09/2022*

- Developed an interactive user interface in **Rshiny** and provided an automated solution to improve the clinical dataflow efficiency
- Spearheaded a cloud storage design with visualizations to improve workflow by 15%
- Socialized and presented the application in alignment with the business objectives to the internal stakeholders

Pathloom — **San Ramon, CA**
Data Engineer — *01/2022 - 03/2022*

- Documented user requirements and developed an **ETL** workflow in **Python** to generate interactive elevation profile and calculate geometry information to provide specifcs for more than 20,000 trails in the US
- Implemented data wrangling in **Python** for exploratory data analysis, outlier treatments and features preprocessing
- Collaborated with product manager to evaluate the performance of cleaned dataset in the testing environment

PROJECT EXPERIENCE

Personalized System Under VAE Framework — *07/2023 - 10/2023*

- Devloped a random utility collaborative model to analyze user choice behaviors within a vast dataset from a smart application
- Employed Python to script a VAE framework and achieved a 5% increase in system performance through the optimization

Data Science for All Fellowship — *06/2023 - 07/2023*

- Selected to participate in 7 week data science fellowship with learning facilitated through real world data science business cases
- Collaborated with group of five to analyze the socio-economic factors that contribute to high school student dropout rates in Los Angeles County using **python, Excel and SQLite**
- Built data pipeline using Python and analyzed five California education datasets with over thousands of county level data
- Created detailed project plan with biweekly meetings to assign deliverables and ensure timeline remained on track, received mentorship and guidance from an industry data scientist
- Present findings to an audience of two hundreds accompanied by a one page data folio using **Tableau**, communicating complex insights to both technical and non-technical audiences

Default of Credit Card Cleints Project — *10/2022 - 12/2022*

- Conducted predictive analysis on customers' default payments using **logistic regression, random forest**, and **XGBoost** classifiers
- Achieved data balance by implementing **SMOTE** technique to oversample the minorty class
- Tuned hyperparameters and applied a 10 fold **cross-validation** model performance, with the XGBoost model performing the best, achieving an 86% F1 score for credible customers class and 50% F1 score for non-credible customers class

Bayes Spatial Model Summer Research Project — *05/2020 - 08/2020*

- Built a generalized latent spatial factor model to model death and treatment rates for 88 counties in Ohio to assist the policy makers decision making in the allocation of resources across the state
- Utilize Nimble framework in **R** with Monto Carlos sampling methods, created **map visualizations** to examine treatment and death rates across the country and selected the best model by evaluating on R squared metrics and credible interval

PERSONAL WEBSITE

- **Github:** https://github.com/luoh17