

Chengyuan Liu

chl4034@med.cornell.edu | (206)604-3027 | <https://www.linkedin.com/in/chengyuan-liu-100698197/>

EDUCATION

Cornell University

Expected Dec 2024

M.S. in Data Science & Biostatistics

- **Courses:** AI in Data, Database Management, Causal Inference, Hierarchical Modeling & Longitudinal Data Analysis

University of Washington

Jun 2023

B.S. in Applied Computational & Mathematical Science (Statistics & Data Science Track) / GPA: 3.8/4.0

- **Courses:** Machine Learning, Stochastic Process, Data Management and Database, Statistical Data Science

TECHNICAL SKILLS

- **Programming:** Python (Pandas, Scikit-Learn, PyTorch), SQL, R
- **Data Tools:** Tableau, Excel, Databricks, Pyspark, Power BI, Git, SAS, GCP, AWS, Snowflakes
- **Data Science Methods:** A/B Testing, Data Wrangling, Database Management, Data Visualization, Time Series Forecasting, Machine Learning, Deep Learning, Statistics

PROFESSIONAL EXPERIENCE

Research Assistant

Nov 2023 – Present

Weill Cornell Medicine

New York City, New York

- Handling datasets with over 1000 samples and 30,000+ variables in application of AlphaFold in microbiome research.
- Employing hierarchical clustering and classification using Scikit-Learn and PyTorch, analyzing high-dimensional data.
- Engaging in the implementation of interpretable learning modules to understand AlphaFold's predictions.
- Conducting extensive literature reviews to latest developments in AlphaFold applications and microbiome data analysis.
- Collaborating in a multidisciplinary team, contributing to discussions and strategies for future research directions.

Data Analyst Intern

Dec 2022 – Aug 2023

Kingdom Consulting

East Windsor, NJ

- Gathered and cleaned user engagement data in **SQL**, conducted EDA, performed statistical analysis, created data visualizations, built and tested **XGBoost & K-means** model in Python to provide prediction of customer behavior, with over 90 % AUC-ROC
- Aligned and tracked key performance indicators (KPIs) with business objectives, facilitating data-driven decision-making and performance monitoring
- Constructed and tested data pipeline and fully automated & interactive **Tableau** dashboards for internal stakeholders, enabling future research on product optimization
- Collaborated closely with cross-functional product and engineering team to perform rigorous data analysis for validating business hypotheses, delivering evidence-based insights and guiding the formulation of strategic recommendations

Data Scientist Intern

Jun 2022 – Sept 2022

Societe Generale

New York City, NY

- Performed SQL queries to build data collection, storage, and processing infrastructure. Analyzed business problems and converted report data to actionable items; delivered simplified and visualized analytical outputs
- Conducted statistical analysis (e.g., Pearson's chi-squared test), built predictive models (e.g., Random Forest, Logistic Regression and SVM) in python to detect fraud, resulting in a 30% increase in detection accuracy
- Designed and implemented ETL solutions with over 200GB data. Maintain data infrastructure including building more than 40 data tables, reconstructed the databased structure, improving the query efficiency about 50%
- Conducted and analyzed 10+ A/B testing experiments to evaluate policy efficiency improvement

Data Analyst Intern

May 2021 – Aug 2021

China Fortune Securities Co Ltd

Beijing, CN

- Analyzed historical share transactions and market fluctuations to simulate risk for share pledge negotiations with a Fortune 500 company in China
- Facilitated collaboration between a private equity quantitative funds, optimizing trading strategies and coordination
- Utilized Python Matplotlib to visually represent financial data for diverse investors, delivering presentations to over 50 investors and establishing a streamlined presentation process

Research Assistant

Jun 2021 - Jan 2022

New York University

New York, NY

- Developed an LSTM algorithm to forecast healthcare sector stock prices, with Min-Max Normalization to compute cell states to mitigate gradient vanishing and explosion, achieving a prediction accuracy of 90% for the next five days
- Designed and implemented a customized stock-picking strategy that significantly enhanced investment performance within the healthcare sector
- Established a robust validation framework to ensure the reliability and resilience of the forecasting model

PUBLICATION

Chengyuan Liu, Eran Wo, Zeyi Chen, *Return Prediction for Healthcare Sector Stocks by Using Long Short-Term Memory Algorithm*, 23 August 2022, <https://doi.org/10.1117/12.2646886>, Proc. SPIE 12330, International Conference on Cyber Security, Artificial Intelligence, and Digital Economy

(2022)