

Alan Wu

Philadelphia, Pennsylvania | +1 (267)-800-3686 | wyulun@seas.upenn.edu | www.linkedin.com/in/yulunwu02

EDUCATION

University of Pennsylvania <i>Master of Science in Data Science</i>	08/2024 – 05/2026 (Expected)
♦ Overall GPA: 4.00/4.00	
University of Nottingham <i>BSc Hons Mathematics with Applied Mathematics</i>	09/2020 – 07/2024
♦ Overall GPA: 3.90/4.00 First Class Provost's Scholarship (06/2021)	

SKILLS

Software Skills: C, C++, Python, MATLAB, R, SQL, PowerBI, Tableau, Spark, Hadoop, Microsoft Office, Jupyter Notebook, AWS

Analytic Skills: Machine Learning (Supervised & Unsupervised), Predictive Modeling, Statistical Modeling, Time Series Analysis, Data Wrangling, ETL, Feature Engineering, Data Cleaning, Data Mining, Experiment Design, A/B Testing, Sample Size and Power Calculation, Data Visualization, Optimization, Hyperparameter Tuning, Model Validation, High Dimensional Model Selection

PROFESSIONAL EXPERIENCES

Novo Nordisk (China) Pharmaceuticals Co., Ltd. <i>Data Analyst Intern</i> Beijing, China	06/2023 – 09/2023
♦ Collaborated with Microsoft China on a company-wide SQL database migration project regarding to customer representative behavior, consolidating five databases into one. Reduced query runtime by 30% through SQL optimization , increasing database migration efficiency	
♦ Leveraged Python and statistical techniques to process and analyze historical sales promotion data for pharmaceuticals. Generated predictive insights that improved promotional budget allocation accuracy by 20%, contributing to \$100,000 in cost savings.	
Central Institute of Mental Health <i>Data Analyst Intern</i> Mannheim, Germany	01/2023 – 05/2023
♦ Explored neurophysiological activity under stress conditions using a neuropsychological paradigm and combining functional MRI and heart rate signal (in collaboration with the Max Planck Institute of Psychiatry)	
♦ Analyzed fMRI-BOLD and photoplethysmogram signals respectively utilizing generalized linear models (GLM) and wabp algorithm in MATLAB and found brain regions' interaction with the autonomic nervous system under stress	
♦ Collaborated with a multidisciplinary team of researchers to prepare and deliver a comprehensive presentation at an academic seminar at Heidelberg University, engaging over 20 scholars and researchers in discussions about findings.	

PROJECTS & RESEARCH EXPERIENCES

News Source Classification	12/2024
♦ Developed binary classification models to predict news headlines' sources (Fox News vs. NBC News), leveraging advanced NLP techniques to reveal stylistic and vocabulary differences.	
♦ Web-scraped and processed 3,800+ headlines using Python , transforming text data via TF-IDF and Word2Vec vectorization, and enriching features through bigram/trigram extraction.	
♦ Built and tuned a Bidirectional LSTM model (80% accuracy), outperforming a baseline Logistic Regression (66%); mitigated overfitting through extensive hyperparameter optimization.	
Customer Churn Analysis	08/2024
♦ Led a comprehensive customer churn analysis using Python , analyzing data from over 7,000 customer records to identify key factors such as contract length, monthly charges and customer service response time influencing churn rates	
♦ Applied the SMOTE method to increase sample size and developed multiple classification models (Logistic Regression, Random Forest, XGBoost) to predict potential customer churn. Achieved an accuracy of 0.9 using XGBoost and created SHAP value charts to explain the model	
♦ Designed and implemented A/B testing to validate data-driven strategies for reducing customer churn, comparing pre- and post-intervention churn rates to measure the effectiveness of the campaign	

Certification

Data Engineering Fundamentals – IBM	01/2025
Professional Certificate in Large Language Model Operations (LLMOps) – AI edX	12/2024
Scrum Fundamentals Certified (SFC) – SCRUMstudy	04/2024
Microsoft Certified: Power BI Data Analyst Associate – Microsoft	11/2023