

Ying Liu

341-345-7248 | ying.liu0331@gmail.com | [Portfolio](#) | [LinkedIn](#)

EDUCATION

Georgetown University

Master of Science in Data Science and Analytics, GPA: 4.0/4.0

Washington DC, US

Aug 2020 – May 2022

Shanghai International Studies University

Shanghai, China

Bachelor of Arts in Business English, GPA: 3.82/4.0

Sep 2016 – Jun 2020

- **Awards:** State-level Merit Graduate (Top 1%), Leadership Excellence Award (Top 5%), Merit Scholarship (Top 5%)

EXPERIENCE

Senior Data Analyst (*Promoted in 2024*)

Jun 2022 – Present

Kaiser Permanente, Financial and Operational Department

Rockville, MD

- **Modeling & Visualization:** Developed forecasting models to predict autism therapy costs, designed interactive features to explore variable impacts, and visualized cost trends in Tableau, achieving a 10% reduction in external costs
- **Financial Analytics:** Collaborated with healthcare stakeholders to refine metrics for data-driven financial reports. Managed ETL processes using cloud-based pipelines (ThriveAI) to streamline data wrangling and accelerate reporting workflows by 20%
- **Database Optimization:** Optimized Microsoft SQL Server database triggers and stored procedures for an internal budgeting tool. Collaborated with SDE team to enhance front-end functionality, improving efficiency and accuracy in staffing allocation

Data Engineer

Sep 2021 – May 2022

Georgetown University: Massive Data Institute

Washington, DC

- **Automated Pipeline Development:** Designed and implemented end-to-end pipelines to automate the storage of images, files, and raw text scraped from websites into a MongoDB database on Docker
- **Data Collection Expansion:** Led a large-scale data collection campaign, gathering data from 100,000 U.S. public schools
- **Web Scraping Tool Release:** Collaborated with a cross-functional team to launch an alpha version of Crawl4All, a web application for automated web scraping, and documented the entire development process

Data Analyst Intern

Mar 2021 – Jul 2021

Didi Chuxing Technology Company, International Business Group

Beijing, China

- **KPI Analysis:** Analyzed 30 key performance indicators from 60 million overseas user trip records, identifying abnormal fluctuations and developing strategies to improve user experience
- **Product Experiments:** Proactively coordinated with product managers to streamline data-related processes in product experiments, including troubleshooting A/B testing, correcting data inconsistencies, and developing new indicators
- **Reporting Automation:** Automated manual reporting processes by creating interactive Tableau reports using Python and SQL, improving efficiency and reducing labor costs by 75%

PROJECTS

Healthcare Provider Fraud Detection

Apr 2024 – Jun 2024

Utilized machine learning to predict potentially fraudulent healthcare providers based on profile and behavioral features

- **Data Preprocessing:** Cleaned and transformed diverse insurance datasets (inpatient, outpatient, beneficiary, claims)
- **Feature Engineering:** Reframed the business question and engineered behavioral features using healthcare domain knowledge
- **Exploratory Data Analysis (EDA):** Conducted in-depth EDA to identify behavioral patterns, such as unusual billing practices and service frequencies
- **Machine Learning:** Trained and fine-tuned models (SVM, Decision Tree, Random Forest, XGBoost) with GridSearchCV, achieving 82% accuracy and 85% recall, with XGBoost outperforming the others

Analyzing Factors for Advertisement Click Rate

Mar 2021 – May 2021

Utilized machine learning and cloud computing to analyze factors that impact the click rate of advertisements

- **Cloud Service Startup:** developed a cloud-based environment for analysis on AWS EMR and PySpark
- **Data Preprocessing:** processed 70 million data points stored in a S3 bucket, cleaned and merged the data tables using spark.sql
- **Feature Engineering:** applied Spark's StringIndexer and OneHotEncoder functions to encode the features including geo-targeting information, transformed the data by VectorAssembler, and established a pipeline to streamline this process
- **Machine Learning:** trained and fine-tuned XGBoost and Random Forest models, achieving an AUC score of 82%

SKILLS

Certifications: AWS Certified Machine Learning - Specialty, Tableau Desktop Specialist

Languages: Python, C/C++, SQL (Postgres/Oracle), JavaScript, HTML/CSS, R, SAS

Tools: Git, Docker, Google Cloud Platform, AWS, Azure, MongoDB, Hadoop, Spark, Power BI, Flask

Libraries: Pandas, NumPy, Matplotlib, TensorFlow, PyTorch

Statistical Analysis: Regression Analysis, Hypothesis Testing, A/B testing

Soft Skills: Communication Skills, Problem-Solving, Teamwork and Collaboration, Attention to Detail, Curiosity and Learning