JEFFREY YUAN

Glenview, IL | mailjeffreyyuan@gmail.com | 847-905-6888 | Personal Portfolio | Research Repository | Github

EDUCATION

Northwestern University, Evanston, IL

MS in Statistics and Data Science June 2026 (expected)

BS in Data Science and Computational and Systems Biology, GPA: 3.98/4.00

Thesis: Deepening Drug Discovery with Causal Inference and Generative AI: Multi-Modal Integration Leveraging Large Language Models and Geometric Deep Learning for Novel Compound Prioritization

- Honors: CME AI/ML Team Challenge Winner, MD+ Datathon Runner Up, Northwestern Summer Undergraduate Research Grant, Molecular Biosciences Summer Grant, Northwestern Conference Travel Grant, and Deans List (8/8)
- Courses: Deep Generative AI, Machine Learning on Graphs, Advanced Machine Learning, Information Management, Data Structures and Algorithms, Statistical Theory and Methods, Data Science with Python, Bioinformatics, Biostatistics

EXPERIENCES

Stats 303-1,2,3 Data Science with Python, *Teaching Assistant*, Evanston, IL September 2024 - Present

- Delivered bi-weekly review sessions on data manipulation, EDA, advanced visualization, and enrichment and held weekly office hours for regression, tree-based, bagging, and boosting machine learning algorithms
- Mentored 5 teams to perform EDA, create visualizations, and build proposal for youth program funding budget in Chicago
- Determined accuracy standards for classification problem attempting to predict on-time order fulfilment of loan applications

CME Group, *Year Round Data Science Intern*, Chicago, IL June 2024 - Present

- Designed automated Agentic AI for natural language to GCP Bigquery (nl2sql) data retrieval for CME Globex order book, Salesforce, Tag50, Marketing Cloud, and Google Analytics databases - achieved a 99% increased productivity
- Productionalized a full-lifecycle MLOps solution on Vertex AI, integrating Feature Store, automated training pipelines, Model Registry, and proactive monitoring for Product Market Qualified Leads (PMQL) prediction
- Trained deep learning models utilizing Commitment of Traders, time and sales, and volatility data to predict weekly returns of 10-year treasury note futures. Determined the efficacy of COT data, resulting in CME publishing COT data on all products
- Developed BigQuery SQL code repository for CME API customers to perform TWAP/VWAP calculations for benchmark products, enable scalability of Data Services product to 100+ customers and achieve $360K in annual revenue

Learngle, *Founding Engineer*, Boston, MA February 2023 - March 2024

- Built RAG-LLM (GPT-4) pipeline for AI/ML technology, clinical case, and clinical implementation question generation and SuperMemo 2 based adaptive learning algorithm for personalized content presentation using a MongoDB database
- Created personalized user performance analytics dashboard, advanced analytics, and performance report with Tableau
- Integrated into the Massachusetts General Hospital, Harvard Medical School, Clinical Informatics fellowship program

Significance Lab, Harvard Medical School, *Research Assistant*, Boston, MA December 2022 - Present

- Fine tuned144 statistical learning models on the MIMIC IV dataset with Azure ML, optimized each model through 5-fold cross-validation and Bayesian optimization. Used predictions and feature SHAP values for enhanced ER resource allocation
- Performed retrospective cohort study on shock index trajectory, categorized patients into 5 groups via clustering algorithms, determined effectiveness with the Bayesian Information Criterion - validated with ANOVA and chi-squared tests
- Leading team of 3 undergrads in performing a web-analysis of Clinical Informatics fellowship pages utilizing the Screaming Frog web-scraping API and coordinated with Massachusetts General Hospital physicians to develope data extraction template

PROJECTS

LLM Bias in Clinical Reasoning, MD+ November 2024

- Recognized medical education bias in tuned LLM (Llama 3.1) utilizing HuggingFace and 10,000+ STEP-1,2, and 3 questions
- Tuned LLM performance using progressive prompt engineering, identifying prompt structures with reduced bias

Airbnb Prediction Problem, Northwestern University January 2024 - May 2024

- Performed exploratory data analysis, feature engineering, and preprocessing on Chicagoland Airbnb data utilizing Python
- Ensembled several boosting and bagging models to 95% accuracy in host status classification and $9.60 in price regression

Deciphering Diabetes, Northwestern University September 2023 - November 2023

- Analyzed Diabetes data from 130 US Hospitals from 1999-2008 to determine the state of Diabetes in the US, demographic and prior medical care contribution to outcomes, and drug efficiency in treating diabetes, utilizing Python
- Presented drug misuse and demographic bias findings to Northwestern Medicine's medical, admin, and business teams

SKILLS

- Technical: Python (Tensorflow, Keras, Pytorch, Scikit-Learn, Pandas, Numpy), R, SQL, NoSQL, Java, Git, Tableau, GCP
- Analytical: AI/ML, deep learning, data science, feature engineering, statistics, big data analytics, probability

INTERESTS

- Competitive swimming, outdoor running, road biking, pick-up basketball, fingerstyle guitar, volunteering