

Shiyu Liu

shiyuliu001@gmail.com • (949)294-0851 • <https://shiyuliu.com>

EDUCATION

M.S. in Data Science, Brown University

To be conferred May 2021

- GPA: 4.0 out of 4.0
- Coursework: Machine Learning, Data Engineering, Statistical Learning, Deep Learning, Computer Vision, Computational Probabilities and Statistics, Hands-on Data Sciences, Data Ethics
- Teaching Assistant for [DATA1010](#) (Math and Probabilities for Machine Learning), [DATA 2040](#) (Deep Learning), [CSCI 1420](#) (Machine Learning)

B.S. in Cognitive Sciences; Statistics Minor, University of California, Irvine

March 2019

- GPA: 3.85 out of 4.0; Award: Magna cum Laude, Dean's Honor List
- Coursework: Multivariable Calculus, Matrix Algebra, Statistical Methods, Multivariate Analysis and Bayesian Analysis

SKILLS

- **Programming Language and Tools:** Proficient in Python, R, Julia, SQL, MongoDB, and SPSS; Basic knowledge of MATLAB, C, Neo4j, Spark and Linux command line; Version control and software product management in Github, Jira, Bitbucket and Confluence.
- **Statistical Analyses and Machine Learning:** GLM, PCA, Factor Analysis, Decision Tree (Boosting) Algorithms, KNN, LDA, SVM, EM, Causal Inference, time series, deep learning with Keras, Tensorflow and Pytorch.

WORKING EXPERIENCES

AI and Data Sciences Intern

[Giving Tech Labs](#), Seattle, WA

December 2020-May 2021

- Conducted research on domain-specific knowledge graph for the public interest; Extract information from authoritative website to match non-profit foundations to donation recipients and generate insights on nonprofit money-flow characteristics using unsupervised learning models.
- Extracted physical features from human voice streams; analyzed and modeled the acoustic measures to predict speakers' age, emotion and healthy condition using supervised machine learning techniques.

Quantitative Research Intern

[Alexandria Technology, Inc.](#)

June 2020-December 2020

- Used NLP techniques and ensemble of different machine learning methods to extract sentiment from news and generate sentiment scores; Extended the multifactor models from US stock market to European and Asian Pacific market.
- Achieved stable 5% active return against US market benchmark and 9% active return against Europe market benchmark on a monthly basis rebalancing strategy for a portfolio around 70 securities; Built interactive visualization and summary tools by economic regimes and sectors.

Machine Learning Engineer Intern

June 2019-August 2019

[Empyrean Software](#), San Jose, CA

- Carried out the development, testing, parameter tuning and optimization of a machine learning product aiming to efficiently predict process corners values in semiconductor wafer based on PVT and corner feature values.
- Utilized neural networks and boosting algorithms to predict the corner values with an average accuracy over 0.97; Improved algorithms and data structures to increase the program running speed 30 times as the original testing speed.

Financial Assurance Intern

June 2018-August 2018

[PwC](#), Beijing, China

- Participated in the interim review of a public Real Estate Group with business coverage over leasing, emporium, hotel operations and multiform investments; Obtained, inspected and analyzed over 200 financial reports, financial agreements, commercial contracts and credentials; Conducted the confirmation procedure for most categories on financial statement; Collected and verified client's statistics through obtaining third-party documents, recalculation and necessary statistical methods.

SELECTED PROJECTS AND RESEARCH EXPERIENCES

[Brown Datathon 2020 Dassault Systèmes Challenge, First Place](#)

February 2020

- Used regression methods, Deep Neural Networks and Boosting algorithms to predict Additive Manufacturing(AM) temperatures and melt dimensions given laser speed, power, direction and edge effects; Achieved accuracy scores over 0.999 for multiple tasks.
- Implemented feature engineering methods based on thermodynamics and PDE, bootstrapping and customized parameter tuning strategies to avoid data leakage problem and improve algorithm stability.

[Sports Data Visualization and Virtual Match Simulation](#)

December 2019

- Obtained NBA match statistics from websites automatically and periodically with web crawler; Visualized, compared and ranked players' performances and capabilities based on users' queries.
- Built data engineering pipeline with Python and MongoDB; Developed web user interface using *Dash by Plotly*; Stored and retrieved data in MongoDB before applying algorithms for virtual match results score calculation.