Baosen Luo

(510)–5705655 | bl3243@nyu.edu | linkedin.com/in/baosenluo

Education

New York University, NY — Sep 2021 – May 2023(Expected)
Master of Science in Data Science — GPA: 4.0 /4.0
University of California, Berkeley, CA — Aug 2017 – May 2021
Bachelor of Arts in Data Science, Economics — GPA: 3.8/4.0

Languages and Skills

- Python, MySQL, Pytorch, Hadoop, Spark, Tableau, Latex, Shell, Git
- Statistical Modeling, Deep Learning, Machine Learning, Visualization
- Convex Optimization, Hypothesis Testing, Time Series Analysis, Big Data,

Related Experiences

Data Scientist Intern | Roblox — San Mateo CA, May 2022 – Aug 2022

- Created a user level metric, Avatar Uniqueness, based on 2d thumbnails using DBSCAN clustering and Resnet-50 embedding; evaluated the metric by survey and correlation analysis; built a data pipeline to monitor the temporal evolution of the metric; proposed actionable product ideas
- Deployed an NLP framework based on Online Latent Dirichlet Allocation to automatically capture trending topics and detect emerging topics from help center tickets; analyzed users' frictions and pain points using the model results; identified a game engine bug and several glitches during the 2-month experimentation

Business Analyst Intern | Tencent Cloud — Palo Alto CA, Apr 2020 – Aug 2020

- Built a pipeline for analyzing and visualizing sales data extracted from Salesforce via Python and Tableau
- Conducted target market analysis and produced market research reports; contributed to Tencent America's 'Go-China' strategy by identifying 500+ potential clients in North America and successfully secured 102 Proof of Concept accounts and 26 Contract accounts

Projects

Movie Recommendation System — March 2022 – May 2022

- Built a collaborative-filtering based movie recommendation system using Apache Spark and high-performance computers; tuned hyper-parameters of Latent Factor Model using cross validation
- Assessed the quality of learned hidden representations of users and movies by visualizing the high dimensional representations in two-dimensional space using UMAP and t-SNE

A Model Comparison of ARIMA and GP in NBA popularity Forecasting — Oct 2021 – Dec 2021

- Performed Box-Jenkins's methodology and cross validation to fit an ARIMA model that captures the underlying trending, seasonal, and periodic structure within the data; achieved RMSE of 6.5
- Designed the kernel function and fine-tuned the parameters for the Gaussian Process Regression that consists of a combination of RBF, ExpiSineSquared, and Rational-Quadratic kernels; achieved RMSE of 5.0

Hypothesis Testing for Movie Ratings — Feb 2020 – Mar 2020

- Applied Welch-T, Mann Whitney, KS, and Bootstrapping tests to study research questions like whether people with different demographics view the same movie differently; are movies that are more popular (operationalized as having more ratings) rated higher than movies that are less popular?