

# Li Sun

Cambridge, MA • 585-957-4337 • [lsun@g.harvard.edu](mailto:lsun@g.harvard.edu) • [LinkedIn](#) • [Github](#)

## EDUCATION

### Harvard University

*Master of Science in Data Science*

**Cambridge, Massachusetts**

*Anticipated Dec 2022*

- GPA: 4.00 out of 4.00
- Related Coursework: Machine Learning, Systems Development, Linear Models, Data Science, Data Visualization

### University of Rochester

*Bachelor of Science in Data Science & Bachelor of Arts in Financial Economics*  
2021

**Rochester, New York**

*May*

- GPA: 3.96 out of 4.00; Phi Beta Kappa Honor Society; Dean's List All Eligible Semesters
- Related Coursework: Data Mining, Database Systems, Computational Statistics, Data Structures & Algorithms, Time Series
- Leadership Experience: Producer of a theatre group (Led a group of 45 students, oversaw and orchestrated four productions with donation increases 10% every semester)

## PROFESSIONAL EXPERIENCES

### University of Rochester Medical Center

*Research Data Analyst (Python, EDA, NLTK, Face++, Ethnicolr)*

**Rochester, New York**

*Dec 2019 – Aug 2021*

- Deployed Latent Dirichlet Allocation (LDA) and sentiment analysis to monitor 7TB Twitter data including users' discussions surrounding mental health concerns during the COVID-19 pandemic and the ban on flavored e-cigarettes.
- Employed deep learning algorithms (Face++ & Ethnicolr) to infer demographic composition of the 3000 target Twitter users.
- Performed time-series analysis to explore the trends of mental health conditions under pandemic in US.
- Used Beautiful Soup to collect available disposable e-cigarettes brands and flavors from online stores, detected the nature and the promotion techniques in disposable e-cigarettes related tweets after FDA banned e-cigarette usage of certain types.
- Published two papers as first author and three papers as second author to make recommendations on effective mechanisms for future communication about the danger of e-cigarette use.

### University of Rochester Visual Intelligence & Social Multimedia Analytics Lab

*Undergraduate Researcher (Python, R, PyTorch)*

**Rochester, New York**

*May 2020 – Aug 2020*

- Quantitatively analyzed short video content-based cultural differences between Douyin and TikTok, discovered different user behaviors and preferences, recommended differentiated commercialization strategies and application feature designs.
- Employed Faster R-CNN and I3D model to perform object and human action detections in 10k videos.
- Conducted statistical analysis, compared the distributional results through data visualizations along the content dimensions of object quantity, object categories, and human action categories.
- Published a paper as first author and gave a presentation on IEEE Big Data Conference with 500 audiences.

### GF Securities Co.LTD

*Investment Banking Analyst Intern (Excel, Wind)*

**Beijing, China**

*Jun 2019 – Jul 2019*

- Performed financial analysis on ten automobile companies, assessed company financial stability, operating efficiency and industry outlook to provide bond holding recommendation to client.
- Conducted macro-economic research on bond market, monitored and analyzed commodity pricing, central bank capital deployment, government bond return movement to provide weekly market update to the team.
- Assisted in pitch-book preparation on one convertible bond issuance, worked independently on government policy implication and track-record section.

## RELEVANT PROJECTS

### Mortality Prediction and Interpretation, Harvard University

**Oct 2021 – Dec 2021**

- Cleaned 15000+ data by changing data type, joining dataset, imputing missing data, and removing outliers.
- Applied PCA to deal with multicollinearity and reduce runtime and complexity for the later training.
- Performed KNN, Logistic Regression, Random Forest, AdaBoost, and Neural Network on data, and determined Random Forest to be the best model based on F1 score (0.803) and AUC (0.861); Interpreted feature importance using SHAP value.

## SKILLS

- **Programming Language:** Python, R, Java, SQL, Linux; Basic: Spark, JavaScript
- **Tools & Frameworks:** PyTorch, TensorFlow, NLTK, OpenCV, Pandas, Scikit-Learn, Git, Excel; Basic: STATA, Tableau

- **Product Skills:** Hypothesis testing, A/B testing, Data modeling/visualization