Shasha Liao

| sliao7@gatech.edu | 404-918-2173 | LinkedIn | Github | Website|
686 Cherry St NW, Atlanta, GA 30313

Skills

- **Languages/Tools**: Python(Pandas, Sklearn, Numpy, Scipy, NLTK, Gensim, Matplotlib, Seaborn), SQL, Matlab, C++, Git, GitHub
- **Domain Knowledge**: Machine Learning, Deep Learning, Statistical Modeling, Data Structure, Algorithms, Mathematics

Education

Georgia Institute of Technology — Atlanta, Georgia
- *Ph.D. in **Mathematics** (minor in **Statistical Analysis**), GPA: 3.87/4.0* — *Jan. 2017 - Expected Dec. 2021*
- *Master in **Computational Science and Engineering**, GPA: 3.60/4.0* — *Aug. 2019 - Dec. 2020*
- *Visiting Student in Mathematics* — *Sept. 2015 - Sept. 2016*

Nankai University — Tianjin, China
- *Post-graduate studies in Mathematics* — *Sept. 2013 - Sept. 2015*

Hunan University — Changsha, China
- *B.S. in Applied Mathematics, GPA: 3.89/4.0* — *Sept. 2009 - July 2013*

Experience

Data Science Intern - NLP | GeniusMesh Corporation | Atlanta, GA — Dec. 2020 - May 2021

Tools: *Python (Pandas, Numpy, NLTK, Gensim, Spacy, Matplotlib, Seaborn, Jupyter Notebook)*

- Built a scraper API from scratch and guided a team of 4 to scrape career experience data of over 1 million users
- Applied Word2Vec, and Machine Learning algorithms (modified K-Means) to classify job titles into 40 classes
- Cooperated with the UI team to generate career path summary and recommendations with visualizations

Selected Projects

- **Loan Agreements Analysis (Python, Sklearn, Pandas, GCP, NLTK, Spacy, Gensim, Numpy, Matplotlib)[link]** — May 2021
  - Studied 3,205 PDF files (10 - 40 pages) from the World Bank's database of loan agreements (1990 - 2019)
  - Extracted key information (borrower, loan amount, currency, closing date, description) using Entity Analysis in Google Cloud Platform
  - Conducted text preprocessing and applied **Word2Vec, PCA**, unsupervised (**K-Means**, 0.47 accuracy) and supervised (**Ensemble of Random Forest, Logistic Regression, SVM, Neural Networks**, 0.71 accuracy) machine learning algorithms to predict the sector
  - Visualized the results using Seaborn and Plotly; recorded a 5-min video with my team to present our results
- **Machine Learning Projects (Python, Pandas, Sklearn, Statsmodels, Numpy, Scipy, Matplotlib)** — Fall 2020
  - **Handwritten Digits Classifier**
    - Analyzed the MNIST Database of 70,000 handwritten digits (0 - 9) with image size 28×28
    - Fine-tuned 5 multi-class classifiers: **KNN, Logistic Regression, SVM, kernel SVM, and Neural Networks** (0.92 average accuracy)
    - Measured and compared the performance of the classifiers using accuracy, precision, recall, F1 score, and confusion matrix
  - **Multiple Linear Regression** [report (11 pages)]
    - Designed a dataset with 1000 instances and 6 attributes
    - Explored 5 common problems: **outliers, high-levarage points, non-linearity, collinearity, heteroscedasticity**
    - Improved the baseline model ($R^2 = 0.245$) to the under truth model ($R^2 = 0.872$) with a high confidence level after 4 steps
- **Credit Card Default Detection(Python, Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn)** — Oct. 2020
  - Performed EDA and data preprocessing on 30,000 customers credit information on 24 features
  - Fine-tuned Machine Learning Classifiers (**Logistic Regression, SVM, KNN, and Random Forest**) to predict credit card default and evaluated them using F1 Scores (highest = 0.47 with Random Forest), and ROC AUC scores (highest = 0.78 with SVM)
- **PhD Thesis: Theoretical and Numerical Analysis on the Stability of Kevin-Stuart's Flow to 2D Euler's Equations** — Aug. 2019 - present

  Tools: *Python(Numpy, Scipy, Sympy, Matplotlib)*
  - Solved a long-standing (54 years) open fluid dynamics problem [reference]
  - Designed a package from scratch for numerical analysis, optimized it for 4 times, and improved the speed by 20 times

Publication

- S. Liao, Z. Lin, H. Zhu, Linear and Nonlinear Stability of Kevin-Stuart Flow to 2D Euler's Equations (70+ pages, in preparation)
- J. Jin, S. Liao, Z. Lin, Nonlinear Modulational Instability of Dispersive PDE Models, *Arch. Ration. Mech. Anal.*, 231:1487-1530, 2019.

Data Science Certificates

- Data Science Boot Camp: Certificate of Leadership, offered by the Erdős Institute — June 2021
- Data Science Boot Camp: Certificate of Completion, offered by the Erdős Institute — June 2021
- Data Science for All: Women's Summit, offered by Correlation One — Oct. 2020
- Sequence Models, offered by Coursera — Jan. 2020
- Machine Learning, offered by Coursera — Nov. 2019
- Neural Networks and Deep Learning, offered by Coursera — Nov. 2019

Teaching Experience

Georgia Institute of Technology | Atlanta, GA — Jan. 2017 - Aug. 2020

- **Graduate TA**: Linear Algebra, Discrete Maths, Multivariable Calculus, Integral Calculus, Finite Mathematics
- **Grader**: Science of Data Science, Probability and Statistics

Honors and Awards

- Recipient of a *Thank a Teacher* certificate for excellence in teaching, Georgia Tech — Aug. 2020
- Recipient of a *Thank a Teacher* certificate for excellence in teaching, Georgia Tech — May 2019
- China Scholarship Council scholarships, Nankai University — Sept. 2015
- The Third Prize Scholarship, Hunan University — Sept. 2012
- National Encouragement Scholarship, Hunan University — Sept. 2011
- National Encouragement Scholarship, Hunan University — Sept. 2010

Selected research talks and posters

- Topic: Nonlinear Modulational Instability of Dispersive PDE Models
  - KUMUNU Conference on PDE, Dynamical Systems, and Applications, Columbia, MO — Apr. 27 - 28, 2019
  - The 37th Southeastern-Atlantic Regional Conference on Differential Equations (SEARCDE), Kennesaw, GA — Oct. 7- 8, 2017
  - KUMUNU Conference on Dynamical Systems and Applications, Lincoln, NE — Apr. 22 - 23, 2017
  - Analysis of Partial Differential Equations using Dynamical Systems Techniques, Boston, MA — June 1-3, 2016
  - KUMU Conference in PDE, Dynamical Systems, and Applications, Columbia, MO — Apr. 23 - 24, 2016
- Topic: On the Stability of the Kevin-Stuart Cat's Eyes Flow
  - SIAM Georgia Tech Student Conference, Atlanta, GA — Mar. 30, 2019

Leadership Experience

- Data science bootcamp group leader at the Erdős Institute — May 2021
- Content Team member of Data Science at Georgia Tech (DSGT) — Jan. 2021 - May 2021
- Math Graduate Teaching Assistant for 7 different math courses at Georgia Tech — Jan. 2017 - Aug. 2020