

Liu Jason Tan

Website: LiuJasonTan.com • LinkedIn: [linkedin.com/in/liujasontan](https://www.linkedin.com/in/liujasontan)

E-mail: liuta@umich.edu • GitHub: <https://github.com/ljtan>

Education

- **Master of Applied Data Science** (*Expected August 2022*) GPA: **4.00** /4.00
University of Michigan – School of Information, Ann Arbor, Michigan
Relevant Courses: *Data Mining, Big Data, Supervised Learning, Unsupervised Learning, Applied NLP, Information Visualization, Social Media Analytics*
- **Bachelor of Science in Information Systems, Cum Laude** (*May 2020*) GPA: **3.64** /4.00
Stony Brook University, Stony Brook, New York

Professional Experience

- **Poisera – Data Analytics Intern** (*June 2021 – August 2021*)
Performed **web scraping** and utilized **APIs** to access public data for analysis, conducted **qualitative research** through interviews with potential customers for additional data to help the company make **informed decisions**, automated **data collection** to update the company database, identified key product insights to inform company road map and improve customer satisfaction
- **Stony Brook University – Senior Computer Specialist** (*October 2017 - May 2020*)
Assisted computer users with technical issues remotely and at workstations by **communicating** to non-technical users, **problem-solving**, and providing excellent **customer service**

Skills

- Programmed in C, HTML, CSS, Java, R, **SQL**, **Spark**, and **Python** (with libraries such as **Numpy**, **Pandas**, **Keras**, TensorFlow, SciKit Learn, Altair, Matplotlib, **Pyspark**, NetworkX, and **NLTK**)
- Constructed models with supervised and unsupervised **machine learning algorithms** such as **deep neural networks**, **classification**, **clustering**, dimensionality reduction, and **regression**
- Familiar with Data Mining concepts such as Apriori, **Hidden Markov Models**, Fourier Theorem, Dynamic Time Warping, **ARIMA**, Reservoir Sampling, Bloom Filter, and Lossy Counting
- Implemented **Natural Language Processing** (NLP) methods such as **Word2Vec**, **WordNet**, **Part-Of-Speech Tagging**, **LSTM**, and **BERT** for sentiment analysis and word-sense disambiguation

Recent Projects

- **My Voice Data Challenge (2021)**
 - Awarded **first place** for the data challenge, using **Natural Language Processing** (NLP) to analyze text message sentiment regarding the Coronavirus which can affect public policy
 - Streamlined the **lemmatization**, text **encoding**, and **hierarchical clustering** (**BERT**) processes, which creates reproducible results and can be scaled to other short-text data
- **S&P 500 Stock Performance Forecasting and Sector Clustering (2022)**
 - Performed **data cleaning** (filling missing values and normalization), **feature engineering**, exploratory data analysis, and data manipulation for machine learning
 - Determined the optimal cluster of stocks using **k-means**, **agglomerative** clustering, and **affinity propagation** by evaluating these methods on three metrics
 - Classified stocks into top 100 and bottom 100 performing stocks using multiple models with accuracy up to 65% by **tuning hyper-parameters**