Liu Jason Tan

Website: LiuJasonTan.com • LinkedIn: linkedin.com/in/liujasontan

Phone: (347) 764-5660 • E-mail: liuta@umich.edu

Education

- **Master of Applied Data Science** *(Expected August 2022)* GPA: **4.00** /4.00

  University of Michigan – School of Information, Ann Arbor, Michigan

- **Bachelor of Science in Information Systems, Cum Laude** *(May 2020)* GPA: **3.64** /4.00

  Stony Brook University, Stony Brook, New York

Experience

- **Poisera – Data Analytics Intern** *(June 2021 – August 2021)*

  Performed **web scraping** and utilized **APIs** to access public data for analysis, conducted interviews with potential customers for additional data, automated **data collection** to update the company database, and identified key product insights to inform company road map

- **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**, **Seaborn**, **Pyspark**, and **NLTK**)
- Developed projects via Microsoft Access, Excel, PowerPoint, Project, Word, Jira, Confluence, **GitHub**, and IDE such as NetBeans, Eclipse, and **Jupyter Notebook**
- Implemented data science methods such as **preprocessing**, exploratory data **analysis**, **big data**, data **mining**, data **modeling**, data **visualization**, and **Natural Language Processing**
- Constructed models with supervised and unsupervised **machine learning algorithms** such as **neural networks**, **classification**, **clustering**, dimensionality reduction**,** and **regression**

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 **lemmatization**, text **encoding,** and **hierarchical clustering** using **BERT**, which creates reproducible results and can be scaled to other short-text data
- **Quicken Loans Data Challenge (2021)**
  - Optimized call data to make predictions on how frequently to call a client, using **Multi-Layer Perceptron** neural network with the **SciKit Learn** Library.
  - Performed data cleaning, **feature engineering**, data splitting to train/test sets, **hyper-parameters tuning** using GridSearchCV, and model interpretation and evaluation.