

Liu Jason Tan

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

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

Work Experience

- **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, performing diagnostics, **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**)
- Proficient in Microsoft Access, Excel, PowerPoint, Project, Word, Jira, Confluence, **GitHub**, and IDE such as NetBeans, Eclipse, and **Jupyter Notebook**
- Experienced in using data science methods such as **preprocessing**, exploratory data **analysis**, **big data**, data **mining**, data **modeling**, data **visualization**, and **Natural Language Processing**
- Knowledgeable with supervised and unsupervised **machine learning algorithms** such as **neural networks**, **classification**, support vector machine, **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 decisions.
 - Wrote an automation script that provides a scalable tool for researchers that reduced the **sentiment labeling** task from hours to minutes.
 - Performed **lemmatization**, text **encoding**, and **hierarchical clustering** using **BERT**, which creates reproducible results and can be scaled to other short-text data.
 - Presented in a **peer-reviewed conference/symposium**, consisting of top natural language processing or medical researchers and judges.
- **Stock Market Prediction (2019)**
 - Analyzed past stock prices to make predictions for future stock prices, using **Long Short-Term Memory (LSTM)** neural network with the **Keras** Library and **linear regression**.
 - Spitted the data into training, testing, and validation sets, **tuned the hyper-parameters** of the model, and evaluated the model to **prevent overfitting**.