Liu Jason Tan

LinkedIn: https://www.linkedin.com/in/liujasontan/ • E-mail: liu.jason.tan1@gmail.com
GitHub: https://github.com/ljtan • Website: https://liujasontan.com/

Education

Master of Applied Data Science — GPA: **4.00** /4.00
University of Michigan – Ann Arbor — *Ann Arbor, Michigan*
Relevant Courses: Data Mining, Supervised Learning, Unsupervised Learning, Cloud Computing, Natural Language Processing, SQL and Databases — August 2022

Bachelor of Science in Information Systems — GPA: **3.64** /4.00
Stony Brook University — *Stony Brook, New York*
Relevant Courses: Object-Oriented Programming, Data Structures, Database Design, Computer Networks, Computer Security — May 2020

Work Experience

Morgan Stanley — *New York, NY*
Associate, Operational Risk Capital Analytics (Full Time) — *August 2022 – Present*

- Created **end-to-end** risk models from scratch by utilizing **PyCharm**, **Jupyter Notebooks**, **Git, and Python** to automate manual (excel) calculations.
- Developed and optimized **natural language processing** and **machine learning** models to enhance operational risk incident quality assurance, reducing workload by over 50%.
- Collaborated with global cross-functional teams to achieve strategic goals, mitigate risk, and deliver robust results that consistently exceeded stakeholders' and regulators' expectations.
- Served as a **subject matter expert** in capital reporting, mentoring junior team members, and ensuring the delivery of high-quality work ahead of deadlines.

Academic Projects

- MyVoice Data Challenge – Received **first place** by leveraging **NLP** techniques to analyze sentiment in text message surveys regarding COVID-19. Automated data cleaning, text encoding, and hierarchical clustering using **BERT** to improve the efficiency of research and generate deeper insights.
- S&P 500 Stock Performance Forecasting – Achieved **62% precision** with a random forest classifier, a substantial improvement over the 20% precision of a dummy classifier. Successfully categorized stocks into top, middle, and bottom tiers using key equity metrics such as price-to-earnings ratio, dividend yield, and volatility.
- Social Media Monitoring - Developed a comprehensive dashboard for real-time **sentiment and topic monitoring** of company discussions. Utilized supervised and unsupervised learning techniques, including **BERT** for emotion classification (e.g., surprise, anger, disgust) and **non-negative matrix factorization** for topic clustering (e.g., account issues, ordering issues, service issues), to gain actionable insights from social media interactions.

Skills and Interests

- Programming languages: C, HTML, CSS, Java, **R**, **SQL**, Spark, and **Python** (with libraries such as Numpy, Pandas, Keras, TensorFlow, SciKit Learn, Altair, Matplotlib, Pyspark, NetworkX, NLTK, and OpenCV)
- Constructed models with supervised and unsupervised machine learning algorithms such as **deep neural networks**, **classification**, clustering, dimensionality reduction, and **regression**.
- Implemented Natural Language Processing (NLP) methodologies such as **Word2Vec**, **WordNet**, Part-Of-Speech tagging, LSTM, and BERT for sentiment analysis and word-sense disambiguation.
- Interests: Cars (especially autonomous vehicles), running, financial markets

Teaching Experience

- Undergraduate Teaching Assistant for Multivariable Calculus
  - Improved student performance by grading hundreds of assignments and providing comprehensive feedback.
  - Elevated student engagement through proactive monitoring of discussion boards and holding office hours.
- Graduate Student Instructor - Being a Data Scientist (Introduction to Data Science)
  - Initiated thought-provoking conversations and engaged student interests through discussion channels.