Jocelyn (Tzu-Jo) Hsu

jocelynhsu.tjh@gmail.com | Linkedin | (+1) 734-546-1271 | Ann Arbor, MI

EDUCATION

University of Michigan | *Ann Arbor, MI* *Aug. 2023 – Apr. 2025*
Master's in Data Science | GPA: 3.86 / 4.0
Relevant Courses: Machine Learning, Large Language Models, Information Retrieval (GSI/TA, Fall 2024)

National Yang Ming Chiao Tung University | *Hsinchu City, Taiwan* *Sep. 2019 - Jun. 2023*
Bachelor of Business Administration in Finance (Arete Honors Program) | GPA: 4.0 / 4.3
Relevant Courses**:** Artificial Intelligence, Linear Algebra, Mathematical Statistics, Data Structure and Algorithms

TECHNICAL AND LANGUAGE SKILLS

ML & Statistical Skills: Predictive Modeling, Deep Learning, Regression Models, Decision Tree Models, A/B Testing Ensemble learning, Unsupervised Learning, Clustering, Experimental Design, Transformer
Programming Languages: **Python** (*Pandas, PyTorch, TensorFlow, Scikit-learn, Matplotlib*), **SQL,** R, C++, JavaScript
Platforms & Tools: AWS (SageMaker, EC2, S3), MySQL, Git, Tableau, Looker Studio, Spark, MongoDB, Linux
Professional Certificate: **AWS Cloud Practitioner** (Amazon Web Services, Jul 2024)

WORK EXPERIENCE

Data Scientist Intern *May. 2024 – Aug. 2024*
Rocket Companies | Largest retail mortgage lender in the U.S. Detroit, MI

- Led the **end-to-end Data Extraction solution** using *AWS Textract* OCR and *LLM* models *Claude Sonnet* to automate payoff document processing that requires manual data collection; achieved **over 97% accuracy** at the document level, ensured the data quality of upstream raw data and **improved the overall productivity by over 30%.**
- Developed a context-aware Workflow Guidance System using Retrieval Augmented Generation (RAG) to tailor personalized requests; **reduced order processing time by 20%**; brainstormed with business partners to define the customized workflow strategies based on local policies.
- Enhanced internal communication by developing a contextual Email Reply Suggestion System. Utilized **QLoRA-finetuned** Microsoft Phi-2 with Salesforce data retrieved from **MSSQL**, optimizing email experience and response efficiency.
- Designed **production-level microservices** to optimize workflows, shadowing and **collaborating with cross-functional stakeholders** to implement centralized targeted solutions.

Data Scientist Intern *Feb. 2023 - Jun. 2023*
Growth Marketing | SEO & Marketing startup Taipei, Taiwan

- Developed a ***GPT4*** Marketing Content Generator, **improving content creation efficiency** while ensuring brand compliance.
- Automated content publishing and feedback analysis for Facebook/Instagram accounts, processing **over 3000 comments** across **8 accounts** with **sentiment analysis and topic extraction**; significantly enhanced marketing campaign understanding.
- Identified and ranked competitive keywords for 4 client websites by analyzing keyword trends utilizing SEO tools and Google ADS Keyword Planner suggestions.

SELECTED PROJECTS

Wikipedia Search Engine | Information Retrieval *Apr. 2024*

- Engineered a robust search engine indexing **1 million** Wikipedia articles, implementing a **LightGBM** deep learning model for ranking optimization. Utilized over 30 extracted features, including network features PageRank and HITS algorithm.
- Designed and implemented an advanced Query Expansion system with unsupervised method *YAKE* and a fine-tuned *BART* for *Named Entity Recognition*, significantly improving search relevance for enhanced user experience.
- Achieved a **27%** increase in Normalized Discounted Cumulative Gain (NDCG@10) over the baseline BM25.

On the Resilience of US ESG Stocks: Evidences from the COVID-19 Market Crashes Topic Research (published), *2023*

- Conducted an event study to analyze the relationship between ESG criteria and the movement of U.S. stock prices during the COVID-19 market crisis, assessing the resilience of companies in the face of extreme market volatility.
- Performed ***Multivariate Regression Analysis*** of cumulative abnormal stock returns on 3,101 US listed firms' ESG score.

Fake News Detection | Natural Language Processing *Oct. 2023*

- Extracted linguistic features, created n-gram and skip-gram word embeddings for 70k news using *NLTK, Word2Vec*.
- Implemented machine learning models *Naive Bayes, LSTM, Decision Tree, AdaBoost*, *BERT* to classify fake news.
- Employed hyperparameter tuning and stacking techniques to optimize the performance, achieving a f1-score of **97%**.