Jane (JINGYAN) JIANG

(412) 773-2913 | jingyanj@alumni.cmu.edu | https://jingyanjiang.github.io/

EDUCATION

Carnegie Mellon University Pittsburgh, PA

- Master of Information Systems Management - Business Intelligence and Data Analytics Aug. 2021 - Dec. 2022
- **GPA**: 3.68/4.0
- **Related Courses**: Machine Learning for Problem-Solving, Deep Learning, Unstructured Data Analytics, Big Data and Large-Scale Computing, Data Structures, Distributed Systems, Object-Oriented Analysis and Design, Database Management

Xiamen University Xiamen, China

- Bachelor of Economics Sep. 2016 - Jul. 2020
- **GPA**: 3.67/4.0 (Rank 5/40)
- **Related Courses:** Mathematical Statistics, Mathematical Analysis, Data Mining, Machine Learning for Economics

Colgate University Hamilton, NY

- Exchange Student, Economics and Computer Science
- **GPA**: 4.0/4.0 Jan. 2019 - May. 2019

Cornell University Ithaca, NY

- Summer Exchange (Development Economics)
- **GPA**: 4.3/4.3 Jun. 2017 - Jul. 2017

SKILLS

Languages: Python, Java, SQL, HTML, JavaScript

Computer Science: PyTorch, Tensorflow, Hadoop, Spark, MongoDB, Tableau, Sklearn, AWS, Docker, Node.js, Object Oriented Programming, Multi-thread Programming

Data Analytics: BERT, GPT, CNN, BiLSTM, Node2Vec, Graph Attention, Logistic Regression, Random Forest, XGBoost, GBDT regression, TF-IDF, KMeans

RESEARCH EXPERIENCE

Bike-Sharing System Project (Carnegie Mellon University-Heinz College) Mar. 2023 - Present

Research Advisor: Woody Zhu (https://sites.google.com/view/woodyzhu)

- Conducted a comprehensive analysis to determine the **causal impact** of station changes, specifically the addition or removal of stations within specific census tracts, on bike demand and supply. Employed data mining techniques to extract insights from bike activity data and leveraged **IPW estimation** to quantify the treatment effect. Findings revealed a positive correlation between the increase in station numbers and higher bike demand and supply
- Explored potential biases in station location selection through **response curve analysis**. Identified that adding stations within census tracts characterized by lower levels of education yielded a more significant increase in bike demand and supply compared to areas with higher education levels
- Future analysis will focus on optimizing the strategy for bike station location to enhance overall effectiveness

Real Estate Market Appreciation Prediction (Capstone Project with **Acram** as Student Consultant) Aug. 2022 - Dec. 2022

- Scrapped the real estate-related news from the Wall Street Journal, New York Times, and Bing News and obtained the unrelated news from AG News. Parsed and cleaned these unstructured data and manually labeled over 9000 pieces of the news into three categories: company expansion, company contraction, and unrelated
- Trained Natural Language Processing Model-**BERT** to classify the news data, reaching an accuracy of 91%, further improved the accuracy by 5% using **GPT** embedding after project ended as GPT 3 came out
- Extracted the location information from the predicted relevant news using **SpaCy** and finally provided a map dashboard to visualize the output across all states by **Tableau**, which helped the investors filter out markets with high potential for appreciation

New-Energy Vehicle Industry Policy Evaluation (Published by Heinz Journal) July. 2019 - Oct. 2022

- Used a high-dimensional fixed-effect model by **STATA** to examine the impact of new-energy vehicle industry policies on new registrations (i.e., sales) of new energy vehicles in each province within the same month
- Collected the industrial policy documents of each province and categorized them into six groups using **BERT**, including provincial plan, pilot city, promotion policy for charging facilities, charging price concessions, convenient access, and monetary subsidies
- Found that the provincial plan could increase the per capita registrations of battery electric vehicles by about 34.3%, these industrial policies have no significant effect on the purchase of plug-in hybrid electric vehicles, and the policy effects are different between private purchase and government procurement

Ministry of Education, Key Laboratory of Econometrics Sep. 2018 - Sep. 2019

- Built online price index by multi-dimensional fixed-effect regression model, overcame the excessive missing value problem, and automatized the update process
- Used Python to crawl Beijing's business districts data via a distributed framework, managed the data with MySQL, adjusted algorithm for other areas

WORK EXPERIENCE

Edison Lab Remote, US

AI Researcher Jun. 2023 - Present

- Mainly focus on researching and optimizing AI algorithms and creating promotional content for social media accounts
- Conducted social network analysis using **TensorFlow**, employing **Graph representation learning with node2vec** and **Graph Attention Model** to identify similar user groups through cosine similarity calculations on user embeddings
- Employed few-shot learning techniques to fine-tune the **GPT-3.5** model by providing task descriptions and some examples, specifically tailored to enhance document accuracy through proofreading and editing

Biomotivate LLC. Pittsburgh, PA

Software Developer Intern Jun. 2022 - Aug. 2022

- Predicted the dropout decision of mandatory treatments for addicted people using sensor data. Developed an **XGBoost** classification model, reaching an F-1 score of 0.80. The predictive output served as a vital metric for medical practitioners to determine a patient's likelihood of continuing with their treatment plan
- Improved the performance of the XGBoost classification model and scaled out to a data size of 32 GB. Retrained the entire model and designed the pipeline with **PySpark**. Reduced the running time of the model significantly by 50%
- Generated videos with desired scenarios from scratch. Applied the **Disco-Diffusion model** in **Python**, which leverages an AI video generating technique called **CLIP-Guided Diffusion**, and tuned the model to generate videos simply based on input text prompts automatically, improving the customer satisfaction on the videos by 20%

PROJECT EXPERIENCE

Entertainment Event Distributed Searching Application Nov. 2022 - Nov. 2022

Distributed System Course Project

- Created a user-friendly native **Android Application** utilizing the **HTTP protocol**, enabling users to search for information of the entertainment events
- Developed a **Java web service** that leverages Ticketmaster API to provide **RESTful API** to the Android application and deployed it to **Heroku** with **Docker**
- Stored crucial user behavioral data, including search keywords, request timestamps, and API latency to **MongoDB** via the web service. Created a web-based dashboard to provide a comprehensive overview of user access patterns and visualize the aggregated metrics, facilitating better analysis of the data

Law Case Search Tool Oct. 2021 - Nov. 2021

Object-Oriented Programming Course Project

- Developed a Java Application with comprehensive functionalities for searching, adding, and modifying law case data
- Employed **TF-IDF** to extract the feature keywords of the document, applied edit distance to infer user's search query intention and correct typos, and increased the search relevance by vectorizing the document with **BERT** using **Tensorflow**
- Used **KD-Tree, Faiss,** and other **Approximate Nearest Neighbor** methods to optimize the performance of the search result ranking service

HONORS & AWARDS

Graduated with Highest Distinction

Issued by Xiamen University · Jun 2020

1st prize in Research Competition on Public Economics and Policies

Issued by Xiamen University · Nov 2018

Qinghan Scholarship (Top 3%)

Issued by Xiamen University · Oct 2018

OTHERS

- **Languages:** English (Professional Proficiency), Chinese (Native), Taiwanese (Native)
- **Interests:** Piano (Level 10 Certificate), Violin, Swimming, Table Tennis