CHEN Yujun

Data Scientist, AI Engineer, and Data Engineer

Osaka, Japan • +81-80-4338-4738 • chingyokukun@gmail.com • linkedin.com/in/cyujun • github.com/xxtyys • xxtyys.github.io

Data Scientist and AI Engineer with over a year of experience in academic and commercial applications. Specializing in multimodal analysis, NLP, ML, and software development, I have improved F1-scores by up to 8.76% in sentiment analysis on social network datasets. Skilled in developing high-accuracy text classification models for Japanese data using Python, PyTorch, NLTK, scikit-learn, and SpaCy. Proficient in Node.js, TypeScript, and MongoDB for full-stack development.

Work Experience

Panasonic R&D, Osaka, Japan Apr 2024 – Jul 2024
Data Scientist

- Collaborated to develop a system analyzing sleep data, enhancing insights into nighttime awakenings and related physiological/environmental factors.
- Contributed to the validation of system workflows on Node-RED, ensuring efficient system performance.

Looking Up Co., Ltd., Tokyo, Japan May 2023 – Mar 2024
Full-stack Developer and AI Engineer

- Collaborated with a team to develop a web application using Node.js, TypeScript, and MongoDB, which analyzed questionnaire feedback to enhance customer business strategy formulation.
- Enhanced a text classification model using NLTK, scikit-learn, and pre-trained SpaCy word embeddings to process complex Japanese survey data, achieving 97.02% accuracy and boosting training efficiency.

Looking Up Co., Ltd., Tokyo, Japan Aug 2022 – Mar 2023
Data Engineer (Internship)

- Developed a ML model using NLTK, scikit-learn, and SpaCy to filter and analyze Japanese survey data, serving as the foundation model for product development.
- Developed APIs for ML models with FastAPI and Flask, creating user manuals for team access.

Selected Relevant Projects

LLaVAC: Fine-tuning LLaVA as a Multimodal Sentiment Classifier Jan 2024

- Collaborated with a researcher to design a prompt for fine-tuning LLaVA, a Large Multimodal Model, to be a multimodal sentiment classifier.
- Outperformed all previous state-of-the art works by up to 8.76% in weighted-F1 score on a social network dataset (MSVA-single).

Multimodal Sentiment Analysis using Multiple Labels from Different Modalities Mar 2023

- Analyzed and visualized the characteristics of the dataset using visualization tools such as Matplotlib, allowing to design a model and find key points before implementation.
- Developed a sentiment analysis model for social media using CLIP, BERT, and RoBERTa, achieving up to 2% increase in F1-score on MVSA-single and MVSA-multiple datasets over previous studies.

Education

Tokyo Institute of Technology, Tokyo, Japan Apr 2021 – Mar 2023
Master of Engineer — Information and Communications Engineering

South China Agricultural University, Guangzhou, China Sep 2015 – Jun 2019
Bachelor of Business Administration — Management Information System

Skills

Technical Skills

- **Programming Language:** Python, C, Java, JavaScript, HTML/CSS, TypeScript
- **ML Toolkits**: PyTorch, Hugging Face, OpenCV, Scikit-learn, Spacy, NLTK
- **Tools & Technologies**: Linux Server, SQL, NoSQL (MongoDB), Docker, Jupyter, GCP, Node-RED

Languages: Chinese (Native), English (Advanced, TOEIC: 830/990), Japanese (Advanced, JLPT-N1)

Certification: IBM Full Stack Software Developer Professional Certificate (Coursera)