Kahei (Althea) Lam

Phone/WeChat：+86 18825645323 | Email：althealam0531@gmail.com

Linkedin: Kahei Lam's Linkedin | Personal Website: Kahei Lam's Website

EDUCATION EXPERIENCES

Cornell University College of Engineering ORIE - Data Analytics track **(One-year Deferment)** 2025.08-2026.06 (Expect)

Cambridge University 2023 Cambridge University Summer Study Program – AI Project Group 2023.07-2023.09

Wuhan University College of Mathematics and Statistics Data Science and Big Data Technology 2020.09-2024.06

- **Honors：** Third-Class Scholarship for Outstanding Students from the School of Mathematics and Statistics, Title of 'Outstanding Student' from Wuhan University, Baosteel Education Foundation Excellent Student Scholarship, Luojia Exchange Excellent Student Scholarship, First-Class Scholarship for Outstanding Students from Hong Kong&Macao&Overseas Chinese, Second Prize in the 'FLTRP·ETIC' National College Student English Writing Competition, Second Prize in the National College Student Mathematics Competition in Hubei Province, H Prize in the American College Student Mathematical Modeling Competition **Work Link**
- **Average Score：** 87.98/100（rank 15%）
- **Relevant Courses：** Machine Learning, Data Structures and Algorithms, Database Technology, Optimization Theory and Algorithms, Python Programming, Introduction to Data Science, Multivariate Statistical Analysis, Statistical Computing, Numerical Analysis, Probability Theory and Mathematical Statistics, Numerical Analysis

INTERNSHIP EXPERIENCES

TikTok *Global E-Commerce – Data Science | Data Analyst Intern* 2024.08-Now

Python, HiveQL, ClickHouse, Aeulous Shanghai，China

- **UK Monthly Report Dataset Development**: This project aims to build a dataset to provide key metric data support for the UK market operations team. The dataset includes logistics key performance indicators such as Longtail Order Rate, Create to Delivery Hours, Trace On-time Rate, and DNR.
  - Specific development tasks: Built a data warehouse using Hive for the monthly report metrics, and optimized query performance with ClickHouse, completing the entire development process from data preprocessing to performance tuning
- **Weekly Report Automation Program Development**: This project aims to assist SEA operations team members in automating the extraction and updating of weekly report data. It implements an automated program that runs every Monday to update key metric data in the weekly report, helping to alleviate the data retrieval burden on team members.
  - Specific development tasks: Used Python to write scripts that automatically extract the required data from multiple dashboards, covering five SEA countries (Singapore, Malaysia, Thailand, Philippines, Indonesia). Additionally, independently collaborated with the Indonesia team to complete specific requirements for the transformation of the weekly report
- **Monthly and Quarterly Report Variance Attribution**: This project aims to write monthly and quarterly reports for the US, UK, and SEA, and to analyze the changing trends in fulfillment timeliness for Lobal and Global Selling in these regions. It involves detailed attribution of factors contributing to both optimization and deterioration
  - Specific tasks: Use SQL to extract data for writing monthly and quarterly reports, and conduct a breakdown analysis of the variances in metrics based on two scenarios: first, the nature of the changes (continuous vs. non-continuous); second, by stakeholders involved (service providers, couriers, buyers, sellers, and products)

Data Wake | *Data Engineering Intern* 2024.04-2024.08

Python, HTML, Jinja, MySQL, Agent, Git, GitLab, Machine Learning Shenzhen，China

- **Weekly Search Volume Forecast**: This project aims to add indicators related to weekly search volume to the advertising domain dataset, helping users understand the search volume for different ASINs
  - Specific development tasks: Build a regression model using historical search volume data from Amazon ABA
- **Underlying Data Infrastructure and Dashboard Replication**: This project aims to acquire data from the Amazon platform and replicate the relevant dashboards.
  - Specific Development Task: Use the Amazon platform's API to pull data and write scripts to regularly fetch product information, including key metrics such as price, reviews, and sales rankings. Utilize JSON and XML parsing libraries to format the retrieved data and replicate dashboards related to advertising and inventory domains.
- **Inspection Agent Configuration**: This project aims to provide an alert function for the customer's store. When the store meets the alert conditions set by the user, alerts will be triggered in two ways: first, by sending alert notifications through the platform; second, by sending alert information via email.
  - Specific Development Task: Quickly master HTML and Jinja template syntax to develop and implement the sales inspection alert Agent. The inspection results received unanimous praise from seed customers. Additionally, continuously follow up on the implementation, rollout, and feedback of the inspection rules to complete the entire project cycle.

PROJECT EXPERIENCES

Aliyun Tianchi - News Recommendation Competition **GitHub Link**

- This project aims to recommend personalized news content to users, enhancing their click-through rates and satisfaction. The specific development tasks include constructing item-based and user-based collaborative filtering models in the multi-channel recall section, as well as designing a cold start recall strategy.

Fine-tuning Common NLP Tasks Based on Transformers GitHub Link

- This project aims to fine-tune Transformer-related models (such as BERT) for common NLP tasks by utilizing datasets from Hugging Face. The tasks involved include: question-answering systems, Quora question pairs, sentence similarity, text classification, token classification, text generation, summarization, and translation.

Application of Model Averaging Methods in Breast Cancer Survival Time Prediction GitHub Link

- This project aims to predict the survival time of breast cancer patients, in which three survival analysis models (Cox proportional hazards model, Weibull model, and accelerated failure time model based on log-normal distribution) and three model averaging methods (Bayesian model averaging, Bagging, and weighted model averaging) were constructed. The effectiveness of weight selection methods (AIC, BIC, Mallows, Jackknife) was also compared

PROFESSIONAL SKILLS

- **Program Language：** Python, SQL, R, Shell, C, MATLAB, Lingo, HTML
- **Web Dev Tools：** PyCharm, VSCode, Linux, RStudio, DataGrip, XCode, MySQLWorkBench
- **Control Version：** Git, GitLab, GitHub
- **Visualization：** Excel, SuperSet, Tableau, FineBI
- **Language: English**（IELTS: 7, CET-6, CET-4）, Mandarin, Cantonese