

# Chun Hei Siu (Charles)

me@chunheisiu.com | Telegram/WeChat chunheisiu | Kowloon, Hong Kong SAR

## EDUCATION

### **Master of Science in Data Science**

#### **University of San Francisco**

San Francisco, CA, USA

Jul 2019 – Jun 2020

- Scholarship recipient, Fall 2019, Spring 2020.

### **Bachelor of Science in Data Science, Minor in Design**

#### **University of San Francisco**

San Francisco, CA, USA

*cum laude*, Aug 2014 – May 2018

#### *Concentration in Computational Analytics*

- USF Provost's Merit Award recipient, 2014 – 2018.
- Dean's Honor Roll, Fall 2014 – Spring 2018.

## WORK EXPERIENCE

### **Lalamove / Huolala**

#### *Senior Data Engineer*

Shenzhen, Guangdong, China

Nov 2023 – Current

#### *Data Engineer*

Jul 2021 – Apr 2023

- Develop and maintain the internal integrated big data development platform, which allows data warehouse engineers to build data pipelines for internal company data aggregation, as well as BI engineers and analysts to make ad-hoc queries and build daily business reports.
- Develop and maintain the internal data platform, which forwards the JSON payload of third-party HTTP requests to internal Kafka clusters for further processing.
- Lead the effort of localizing various internal data products co-developed by mainland China colleagues, for colleagues based in Hong Kong and various other locations.
- Led the cloud migration of the internal integrated big data development platform from AWS to Azure.

### **IBM**

#### *Junior Data Scientist*

Lantau, Hong Kong SAR

Nov 2020 – Jun 2021

- Designed and implemented data parsing and transformation scripts in Python for Apache Airflow workflows, including XML and JSON to CSV conversion scripts, which are used for data ingestion in cloud databases.
- Developed and integrated Informatica ETL workflows and mappings for business layers in the cloud data warehouse, which simplify OLAP queries for analytics and optimizations.

### **Hong Kong Applied Science and Technology Research Institute (ASTRI)**

#### *Engineer, Software Development*

Shatin, Hong Kong SAR

Aug 2018 – Jun 2019

- Designed and developed REST APIs in the backend module for the Intelligent Ad Management Platform.
- Implemented DevOps such as CI/CD with Jenkins and GitLab CI in the development environment. Improved productivity by eliminating the need for manual image building and deployment.
- Integrated service mesh management with Istio on Kubernetes and presented the topic to team members.
- Designed Kibana dashboards for metric aggregation and distributed tracing in the development environment.
- Maintained workstations and server equipment for software development and client demonstration. Coordinated with the project manager for equipment procurement.

## INTERSHIP EXPERIENCE

### **Valimail**

#### *Data Science Intern*

San Francisco, CA, USA

Oct 2019 – Jun 2020

- Tackled the problem of classifying a backlog of 100K+ unknown internet domains generated by Valimail Defend.
- Developed an end-to-end machine learning pipeline that classifies domains by detecting whether they belong to low-risk categories such as real estate. The Gradient Boosting Machine (GBM) model achieved a 95%+ precision rate with test data when classifying real estate domains using Natural Language Processing (NLP) for web content analysis. Going into production, the pipeline greatly reduced the reliance on manual data labeling.
- Designed and implemented RPC APIs using Flask in Dockerized modules in the pipeline.
- Built web scrapers using BeautifulSoup to gather multiple external data sources for ML model training.

### **Macroview Telecom Limited**

#### *Summer Intern, Software Development*

Quarry Bay, Hong Kong SAR

Jun 2016 – Aug 2016

- Performed quality control tests for in-house software applications integrated with Cisco UC Solutions.
- Developed and presented a demo application integrated with Cisco Jabber Web SDK.

## TECHNICAL SKILLS

- Programming Languages: Java, Python, R, JavaScript, SQL, HTML, CSS
- Tools & Platforms: Hive, Spring Boot, Apache Spark, Docker, Kubernetes, Jenkins, Elasticsearch, Kibana, Redis, Apache Kafka, Apache Airflow, Jupyter Notebook, Informatica, RStudio