

# Chenglin(Celine) Han

celinehan.hcl@gmail.com | celinehan\_hcl@163.com  
https://celinehan-hcl.com/

## SUMMARY

---

Celine is a recent graduate with a Master's Degree in Statistics and hands-on experience in machine learning, statistical modeling, and data analysis. In her last internship, she used advanced ML algorithms to predict the completion of aircraft turnaround activities. In her free time, she tutors students in Statistics at both undergrad and grad levels, sings a lot, and keeps up with the latest trends in the K-pop industry.

## INTERNSHIP EXPERIENCE

---

- Changi Airport Group, Airport Operations** Jul 2023 - Dec 2023  
Data Science and Machine Learning Intern, Sensor ML Lab Singapore
- Analyzed business needs into technical solutions with cross-functional teams of domain experts and project managers to prepare for feature engineering.
  - Collected and preprocessed 20k+ rows of offline data with data science libraries in Python (NumPy, Pandas, Seaborn, etc.), delivering a detailed exploratory data analysis report.
  - Developed and fine-tuned new machine learning models using Scikit-Learn API of XGBoost in Python, achieving an accuracy **competitive against the only ready commercialized solution in the industry**.
  - Deployed the final model in a CI/CD pipeline on the AWS cloud platform as a microservice, delivering an end-to-end machine learning project.

## PROJECT EXPERIENCE

---

### Prediction and Classification of Flight Cancellation and Delay

- Collected and processed 1.5M+ flight records using HQL within the Hadoop framework.
- Implemented logistic regression with PySpark, achieving 99.92% accuracy in predicting flight cancellations and delays and classifying influential factors.

### Forecasting Retail Sales Index of Singapore

- Collected, processed, and visualized monthly retail sales data using R libraries including dplyr, ggplot2, etc.
- Developed predictive statistical models ARIMA and ETS with R, achieving a mean absolute percentage error (MAPE) of 1.09% in forecasting the sales index 12 months forward.

### Analysis of Trending Topics on Youtube

- Processed and aggregated 40k+ YouTube videos using the MapReduce framework in Hadoop.
- Conducted statistical analysis and visualization with Python, providing insights into the most popular 12 content categories and summarizing prevalent video trends in a detailed report.

## TECHNICAL SKILLS

---

- Machine Learning:** Decision Trees, Random Forests, XGBoost, kNN, K-Means, PCA
- Statistical Models:** Linear Regression, GLMs, ANOVA, ARIMA, ETS, Survival Analysis
- Programming:** Python, R, SQL (PostgreSQL), Shell, Git (Bitbucket)
- Big Data:** Hadoop (HDFS, MapReduce), Hive (HQL), Spark (PySpark, SparkR)
- Software:** Microsoft Power BI, Gephi, SPSS

## EDUCATION

---

**National University of Singapore** Aug 2022 - Jan 2024  
Master of Science in Statistics Singapore

GPA: 4.94/5.0 (top 3%)

*Relevant Coursework:* Statistical Learning, Survival Analysis, Regression, Time Series, Multivariate Statistics

**Beihang University (BUAA)** Sep 2017 - Jun 2021  
Bachelor of Economic Statistics Beijing

*Honorary awards:* Outstanding Graduate, Student Leader, Scholarship for Literature and Art

## OTHERS

---

- Languages:** Mandarin Chinese (native), English (IELTS 7.5, professional proficiency), Korean (beginner)
- Leadership:** Co-Chair of the Department of Liaison and Fundraising, SAST of Beihang University
- Interests:** K-pop, Singing (Champion in the Campus Singing Competition of Foreign Language Songs)