

# 韩铖霖

celinehan\_hcl@163.com | celinhan.hcl@gmail.com  
https://celinehan-hcl.com/  
寻找数据分析师/科学家的全职工作机会



## 实习经历

- 樟宜机场集团 (Changi Airport Group)** 2023年07月 - 2023年12月  
数据科学与机器学习实习生 机场运营部 新加坡
- 确认业务需求**：与各部门进行讨论和沟通，定义数据的使用场景，确认预测任务及目标。
  - 数据处理与探索**：使用Python的NumPy、Pandas、Seaborn等模块，收集、预处理和可视化20k+条离线数据，产出详尽的探索性分析报告，并完成特征工程。
  - 模型开发及测试**：使用XGBoost模块中的Scikit-Learn API，建立统计与机器学习模型，对飞机周转活动完成时间进行预测，在测试集上的表现**与业内唯一的商业化解决方案效果相当**。
  - 云端部署**：把模型部署在AWS云平台服务的CI/CD管道中，完成一个端到端机器学习项目的交付。

## 项目经历

- 新加坡每月零售销售总额预测**
- 数据收集与可视化**：收集新加坡1985年1月至2019年12月的月度零售总额指数并导入R，使用dplyr和ggplot2库对数据进行可视化，探索数据的整体趋势和季节特征。
  - 模型拟合与预测**：使用tsibble和fable库中的ARIMA和ETS模型对历史数据进行自动化拟合，并预测未来12个月的零售总额指数，预测误差 (MAPE) 约为1.09%。

### 航班取消和延误的分类

- 数据收集与处理**：通过在shell中执行的hql脚本以及使用Hive SQL中创建外部表操作，将1M+条航班数据导入Hadoop分布式文件系统(HDFS)，并利用Python中的Pandas和PySpark相关模块完成数据的预处理和特征工程。
- 模型拟合与预测**：利用PySpark创建LogisticRegression实例，在训练集上航班取消和延误情况及其影响因素拟合逻辑回归模型，分析导致飞机取消或延误的因素；在测试集上的预测准确率达到99.92%。

### 社交媒体热门视频探索性数据分析

- 数据导入与聚合**：在Hadoop的MapReduce框架下，利用Python和shell脚本，导入与聚合40k+条热门视频数据。
- 可视化与探索性分析**：利用Python的NumPy、Pandas和Matplotlib库对数据进行可视化与描述性统计分析，分析视频数量和点赞的分布，总结出最受欢迎和创作最多的视频类型，并解释热门视频现象背后的原因。

## 专业技能

- 机器学习算法**：决策树、随机森林、XGBoost、kNN、k-means、PCA等
- 统计学模型**：线性回归、广义线性模型、方差分析、ARIMA、指数平滑、生存分析等
- 编程**：Python, R, SQL ( PostgreSQL ), Shell, Git ( Bitbucket )
- 大数据**：Hadoop ( HDFS, MapReduce ), Hive ( HQL ), Spark ( PySpark, SparkR )
- 软件**：Microsoft Power BI, Gephi, SPSS

## 教育经历

**新加坡国立大学** 2022年08月 - 2024年01月  
统计学 硕士 统计与数据科学系 新加坡

CAP: 4.94/5.0 (前3%)

相关课程：多元统计、类别变量、有限样本抽样、空间统计、回归分析、时间序列、生存分析、统计学习

**北京航空航天大学** 2017年09月 - 2021年06月  
经济统计学 本科 经济管理学院 北京

荣誉奖项：校级优秀毕业生、优秀学生干部、文艺优胜奖学金

## 其他

- 语言**：普通话 ( 母语 ), 英语 ( 雅思7.5, CET-6 ), 韩语 ( 初学 )
- 兴趣爱好**：唱歌 ( 外语歌曲大赛冠军 ), K-pop
- 课外活动**：北航大学生科学技术协会 ( 外联部副部长 ), 文艺委员 ( 连续两年组织班级文艺汇演 )
- 经历**：概率统计课程助教, 雅思英语老师