Yiran Jing

Data Scientist and Engineer Intern at Douugh | Undergraduate Student at University of Sydney

yjin5856@uni.sydney.edu.au | 0401246938 | linkedin.com/in/YiranJing | github.com/YiranJing

EXPERIENCE

Data Science and Engineer Intern
Douugh

08/2019 – Present *Sydney, Australia*

An Australian fintech start-up, AI led technology platform

- NLP based analysis
- Using mlflow to manage ML models lifecycle

Data Science Intern
Taysols

06/2019 – 08/2019 *Sydney, Australia*

Australia's leading cloud, consulting for Business Analytics solutions.

- Provision of data cleaning, EDA, feature engineering, hypothesis testing, statistical analysis and ML model turning (multiclass XGBoost, ARIMA, and DeepAR+) services utilizing Python.
- AWS SageMaker for training, deploy and validate standard AWS ML models (hosting service and batch transformation). Turning the best machine learning models utilizing WAS SageMaker customized SciKit models, with business idea/insight suggestion, written up to Latex documents.
- Build Lambda functions for real-time predictions and batch predictions, also clean data within Lambda Function. Monitoring CloudWatch for in-time performance. Give insights about the model performance and further improvements.

Data Analyst
Chain noodle Restaurant Macaroni of Zhang grandpa

07/2019 – Present

Market and customer analysis for chain noodle restaurants

- Build MeiTuan AppSpider to scrape information customer comments from thousands of food restaurants using Python. Then build AWS lambda to automatically update customer comments monthly.
- Using wordcloud, statistical inference, data visualization to analysis the relationship among price, order quantity, eating preference, consumption level, regional price difference, free wifi service, etc.

Project Participant
ANOVA Project

02/2018 – 12/2018 *Sydney, Australia*

ANOVA Project is the first pro-bono STEM Consulting Student Organisation

- Intellify Project (2018 Semester 2:): Topic: Pricing Optimization; Work with ANOVA group members, supervised by Intellify Pty Ltd. Taking into account for cross-product and external effects. Time series, statistics, R and Python knowledge are applied to this project.
- Equitise Project (2018 Semester 1): Data service for Equitise Pty Ltd (Sydney). Provided valuable insights and suggestions based on their customer/trading data. Applied Python and statistical modeling knowledge in this project.

SKILLS

Python; R; Java; SQL; Hadoop; Flink; JavaScript

Modelling: Statistical; ML; (Un)supervised Learning

AWS Lambda; Amazon SageMaker; Amazon Forecast

PROJECTS

Amazon Github (awslabs) contribution for sagemaker machine learning customer churn (07/2019 – Present)

- add batch transform demo for xgboost customer churn, also add customised model demo for customer churn example.

Power Demand Forecast of South Australia (09/2018 – 11/2018)

- Reducing Power Supply Costs in South Australia using Statistical time series and Neural Network, based on 30 mins time series demand data and Bureau of Meteorology Adelaide weather data.

EDUCATION

Bachelor of Science and Advanced Study
University of Sydney

07/2016 – 11/2020

Major Statistics and Computer Science; Minor Mathematics and Business Analytics

- WAM HD +
- Statistics (89); Computer Science (89); Mathematics (84); Business Analytics (92);

NOTABLE AWARDS

Academic Merit Prize and Dean's List for Academic Achievement

University of Sydney

- Awarded to top 600 students every year

Course Rankings Top 3

University of Sydney

- Statistical Tests (STAT2012); Introduction to programming (INFO1103); Analysis (MATH2023); Management Science (QBUS2310); Predictive Analytics (QBUS2820); Advanced Analytics (QBUS3830)

RESEARCH

Undergraduate Research Assistant at Tsinghua University (12/2017 – 02/2018)

Topic: Predict the primary market of Art in China (Econometrics Program). Web-Crawler using python, crawling the resumes of artists automatically on gallery websites. Wrote package utilizing python to automatic fetch keywords in resumes.

Mathematics Research Assistant Summer Scholarship at the University of Melbourne (01/2017 – 02/2017)

Topic: Improve HAR Models for Realized Co-variance: Long-memory Forecasting with Dynamic Attenuation in Multivariate Cases (MATLAB). Applied models to forecast the motion of the realized volatility based on five minutes returns real financial data.