Pei Yingchi

Resume

149 Mei Ling Street, Singapore, 140149
+65 86699823
yingchi.pei@gmail.com
www.peiyingchi.com

About me

- 4 years of experience in data science and machine learning field.
- Experience of building production-level data science solutions with big data environment and cloud services. Familiarity and hands-on working experience with classical ML methods like Logistic Regression, Random Forest, Boosting, NLP techniques and Neural networks.
- Thrive to be a full-stack data scientist who can conduct rigorous research to uncover business insights and also deliver production-level code to build practical applications.
- Keen to learn, experience and share - https://blog.peiyingchi.com

Education

2018 – 2020 **Computer Science (Master)**, *National University of Singapore (NUS)*.
- CAP: 4.85/5.0

2013 – 2017 **Business Analytics (Bachelor)**, *NUS*.
- CAP: 4.91/5.0, Honours with Highest Distinction
- Winner of Lee Kuan Yew Gold Medal (Top student of the batch)
- Five times Dean's Lister

Work

2019 Apr-Present **Data Scientist**, *Indeed.com Ltd.*, Singapore.
Work in salary estimation and career experience team.
- Applied NLP techniques (entity embeddings) with tree-based ML models to estimate job salary using structured as well as unstructured text features.
- Designed and developed the jobseeker salary inference pipeline including model (re)training with AWS SageMaker, model deployment by setting up REST and gRPC service from Python, and model monitoring with scheduled jobs.
- Built Python modules for text summarization and ranking to provide representative content to display on webpages, using NLP techniques such as TextRank and Word2Vec.
- Prototype exploration and exploitation pipeline for dynamic ranking.

2018 Aug - 2019 Jan **Data Scientist**, *Bitmain Technologies Ltd.*, Singapore.
- Predicted blockchain transaction fee using Neural Networks and Generalized Linear Models; building the end-to-end process from acquiring real-time data (Python parser with Redis and MySQL) to training and evaluating models.
- Generated DS insights newsletter for the mining platforms using Airflow scheduled Spark jobs.

2017 Jul - 2018 Jul **Data Scientist**, *DataSpark, Singtel,* Singapore.
- Researched on footfall analytics with telco data using machine learning algorithms such as Naive Bayes, Logistic Regression, and Random Forests. Submitted two research papers based on the research with one published.
- Implemented and productionized models into our data analytics solutions using Python.
- Designed and developed the network planning application for telco operators to reduce upgrading cost while improving customer experience. The application was built with Scala and deployed in a big-data environment with Hadoop and Spark.

2016 May-Nov **Data Scientist**, *DataSpark, Singtel,* Singapore.
Internship.
- Conducted geolocation data analysis projects by running Hadoop and Spark jobs; implemented reproducible code using R Markdown and Python for the projects.
- Built interactive data visualizations (Web apps) using JavaScript, Node.js and React.

2015 **Market Research Analyst**, *Millward Brown*, Shanghai, China.
May-Aug Internship.
- Prepared Budweiser's 2015 Q1 report which was well received by the client.
- Collected and complied the consumer survey data weekly using SPSS Survey Reporter.

Publications

2017 Ng, Y., **Pei, Y.**, Jin, Y. (2017).
"Footfall Count Estimation Techniques Using Mobile Data", 2017 IEEE 18th International Conference on Mobile Data Management (MDM), Daejeon, South Korea, 2017, pp. 307-314. doi: 10.1109/MDM.2017.49

Projects

2015 **Bankruptcy Prediction Using 10-K Reports**, *Bachelor Capstone Project*.
- Gathered data from 3 large financial databases; investigated different methods such as resampling and cost-sensitive learning to deal with the imbalanced dataset.
- Incorporated text-mining features such as fog readability and sentiment score to the traditional prediction model. The best model achieves testing AUC score 0.9.

2014 **Twitter Follower Analysis For "@TopShop"**, *Data Mining Project*.
- Collected data of 20,000 Topshop followers using R through the Twitter API.
- A classifier was built to classify tweets into categories. Compared various methods such as Logistic Regression, Random Forests, SVM, and K-Nearest Neighbours.

Activities

2015 **Teaching Assistant**, *IS1112 E-Business Essentials, NUS*.

2014 **Program Organizer**, *Egyptian Society for Integrated Development*, Cairo, Egypt.
Summer
- Planned and conducted activities such as lectures and field tours for underprivileged children.
- Survived a bomb attack.

2013 – **Vice President**, *Chinese Society Chinese Orchestra, NUS*.
2014
- Organized annual music concert, welcome events, music talks and cultural exchange programs.

Additional Information

Languages Fluent in English and Mandarin. Basic German (A1 level)

Tech stack Python (primary), Scala, R, SQL and SAS; Spark, AWS, Airflow, Docker

Interests Music: Play the piano, the cucurbit flute and the Erhu

Sports: Taekwondo Black Belt and Karate