Pei Yingchi

Resume

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

Education

2018 – 2020 **Computer Science (Master)**, *National University of Singapore (NUS)*.

2013 – 2017 **Business Analytics (Bachelor)**, *National University of Singapore (NUS)*.

- CAP: 4.91/5.0, Honours with Highest Distinction
- Winner of Lee Kuan Yew Gold Medal (Highest CAP of the batch)
- Five times Dean's Lister (Top 5%)

Work

2018 Aug-Present **Data Scientist**, *Bitmain Technologies Ltd.*, Singapore.
Work in btc.com's data team.

- Serve the statistics needed by online mining pool platforms and blockchain explorers using Airflow scheduled Spark jobs, rendering Hive tables through Tableau server dashboards
- Develop a transaction fee prediction engine using Neural Networks and Generalized Linear Models; in charge of the end-to-end process from acquiring real-time data (Python parser with Redis and MySQL) to training and evaluating the models
- Support other business units by providing reports generated using Spark SQL and graph databases like neo4j

2017 Jul-2018 Jul **Data Scientist**, *DataSpark, Singtel*, Singapore.
Work in the application team.

- Researched on footfall analytics with telco data using machine learning algorithms such as Naive Bayes, Logistic Regression, Support Vector Machine (SVM) and Random Forests. Implemented and productionized those algorithms into our data analytics platform using Python
- Developed intellectual property for the company by submitting research papers.
- Designed and develop 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 December **Data Analytics Intern**, *ViSenze*, Singapore.
Fully in charge of the internal dashboards.

- Established the pipeline of internal metrics reporting by understanding the requirements from various team leaders; Assisted in MySQL database design
- Produced dashboards on system and business performance to enable stakeholders to make effective decisions, using Chartio and SQL

2016 May-Nov **Data Science Intern**, *DataSpark, Singtel*, Singapore.
Part of the product team at DataSpark.

- Worked in a big data environment with Hadoop and Apache Spark
- Conducted several geolocation data analysis projects which added new features to our data solutions and improved the model accuracy
- Implemented reproducible code using R Markdown and Python for the projects
- Built interactive data visualizations (Web apps) using JavaScript, Node.js and React for internal and external clients

2015 May-Aug **Market Research Analysis Intern**, *Millward Brown*, Shanghai, China.
Work on a long-term marketing research project for Budweiser.

- Successfully prepared the 2015 Q1 report for Budweiser with another team member in time, which was well received by the client
- Collected and complied the consumer survey data weekly using SPSS Survey Reporter
- Detected several problems and initiated deep dive research to find explanations

Publications

2018 **Pei, Y.**, Ng, Y., Cai, Y., Li, Y. (2018).
"Geospatial Re-Distribution Methods for Footfall Count Estimation using Mobile Network Data", Submitted for Publication

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 1st sem **Bankruptcy Prediction Using 10-K Reports**, *BA Capstone Project*.

- Gathered data with over 50,000 records from 3 large financial databases
- The model assumptions and data consistency were carefully evaluated
- 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 2nd sem **Twitter Follower Analysis For "@TopShop"**, *Data Mining Project*.

- The data of 20,000 Topshop followers were collected using R through the Twitter API
- A classifier was built to classify tweets into categories. Various methods such as Logistic Regression, Random Forests, SVM, and K-Nearest Neighbour were used
- The followers' account attributes were analyzed, and the results were aggregated and visualized using R and Tableau

Activities

2014 – 2015 **Teaching Assistant**, *IS1112 E-Business Essentials, NUS*.

- Crafted and compiled 4 question sets containing over 100 questions
- Provided assistance and feedback for the students throughout the course

2014 Summer **Program Assistant**, *Egyptian Society for Integrated Development*, Cairo, Egypt.
Design and implement activities for underprivileged children.

- Survived a bomb attack
- Gave lectures using audio-visual aids, role-playing and interactive games, resulting a 40% increase in the attendance rate. Organized a field tour about saving water

2013 – 2014 **Vice President**, *Chinese Society Chinese Orchestra, NUS*.

- Organized an annual welcome event, which resulted in a membership increase of 50%
- Expanded the influence of CSCO by organizing an annual music concert with more than 200 audiences
- Initiated 4 new events with the committee, including music talks and cultural exchange programs

Additional Information

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

Skills Statistics background with 12 statistics modules taken. Programming background with knowledge in Python, Scala, R, Spark, JavaScript, SQL and SAS

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

Sports: Taekwondo and Karate