XIAOCHI LI

george_li@gwmail.gwu.edu | (571)225-1934 2091 Briarwood Dr. Santa Clara, CA
LinkedIn: https://www.linkedin.com/in/xiaochi-li/ GitHub: https://xc-li.github.io/

EDUCATION

The George Washington University Washington, D.C., USA
Master of Science in Data Science, GPA:4.0 / 4.0 05/2019
Course: Machine Learning, Deep Learning, Design and Analysis of Algorithms, Natural Language Processing, Bayesian Methods, High Performance Computing and Parallel Computing

East China University of Science & Technology (ECUST) Shanghai, China
Bachelor of Science in Economics, GPA: 3.54 / 4.0 07/2017

TECHNICAL SKILLS

Python (Pandas, Scikit Learn, TensorFlow, NLTK), R, Spark, Tableau, Shell, AWS, SQL

DATA SCIENCE PROJECTS (Read detail on: https://xc-li.github.io/)

Machine Learning: Loan Default Prediction 04/2018 - 05/2018

- Led a 3-student group to build an end-to-end machine learning pipeline based on 887K Lending Club data, and achieved 70% recall score on loan default prediction
- Utilized feature engineering, over-sampling and fine-tuned various supervised models such as Random Forest, Logistic Regression with Scikit Learn to optimize performance

Deep Learning: Facial Expression Recognition 11/2018 - 12/2018

- Managed a 3-student group to classify 12K facial expression images into 7 categories by developing a convolutional neural network (CNN) with TensorFlow and Keras on AWS, achieved 74% accuracy
- Built a real-time facial expression recognition program with OpenCV3 and highly complimented by the professor

WORK EXPERIENCE

FiscalNote Washington, D.C., USA
Data Scientist Intern (NLP) 03/2019 - 05/2019

- Extracted relevant speech from over 40K XML congressional records, developed regular expression rules, evaluated several machine learning models (Random Forest, XGBoost, Logistic Regression) to build a two-stage model for stance detection and stance classification, and achieved 90% F1 score
- Developed a machine learning pipeline with modular design to reduce duplicated code by 50% and improve the speed of text processing by 4 times with parallelization
- Encapsulated the pipeline as a package and wrote unit-test and documentation for deployment

Michelin (China) Investment Co., Ltd. Shanghai, China
Business Intelligence Intern 07/2016 - 02/2017

- Designed a program using Python's Regular Expression to extract data information of tires, drastically reducing the work period from 10 weeks to 2-3 days
- Conducted data collection and analytical research to establish a company wide database for predicting sales figures and market shares of different types of tires