Sophie Chen

Website Sophie-Chen-Data SophieGarden
217-607-4503 Seattle WA
SophieChenPhD@gmail.com

Skills

Programming: Python, SQL, Tableau, Git, Bash, PyTorch, Keras, Scikit-learn, SciPy, MapReduce, Jupyter, Pandas, Numpy, Matplotlib, Matlab, LaTeX, (Basic: Java, Flask, HTML, CSS)

Machine Learning: Supervised Learning, Deep Learning, Image Classification, Unsupervised Learning, Evaluation and Diagnostics

Statistics: Hypothesis Tests, A/B testing, Propensity Models, Time Series Analysis

Experience

Data Scientist Fellow *Insight Data Science, 2018*

- Developed a Chrome Extension using ***customized hierarchical classification*** algorithm with natural language processing (***NLP***) to predict best folder for downloading files
- Achieved 82% predictions on the right path, of which 69% within 1 level away
- Deployed with Python, Flask, Chrome API, PyInstaller (***strong programming skills***)

Optical Modeling Research Assistant *University of Illinois, Fudan, 2010-2018*

- Build ***predictive*** thermal model with ***3D visualization***, improved fiber laser power by 10%+, and broadband fiber source power to ***the highest to date*** (Matlab)
- Verified three-body formation theory for the first time via ***automated data collection***
- Analyzed ***transformed*** spectra and ***time series data*** with computational models, reduced RMSE of molecular potential curves by 20%+, discovered 11+ new electronic transitions
- Developed ***playbook-style documentation*** for managing various lab tasks

Consultant for Renter Evictions *Consulting, 2018*

- Analyzed racial bias in renter evictions with ***propensity modeling***, reported with ***interactive visualization*** (Tableau)
- Achieved 0.7 RMSE on predictions of eviction rate percentages, required extensive ***cleaning and engineering*** of high-dimensional data (XGBoost, Cross Validation)

Projects

Customer Identification for Mail-Order Sales Company

- Identified core customers by ***interpreting the differences between the clusters*** for the general population and that of customers (PCA, K-Means)
- Built ***pipeline*** that includes data cleaning, feature engineering, modeling, clustering

Flower Species Image Classifier

- Designed a ***neural network*** image classifier with ***PyTorch*** (ResNet, VGG)
- Built a Command Line App to automate extract, load, transform (***ETL***), learning, and inference

Education

PhD, *Electrical and Computer Engineering* *University of Illinois Urbana-Champaign, 2018*

Nano Degree, *Data Science* *Udacity, 2018*

MS, BS, *Optical Engineering* *Fudan University, 2013*