Zhaoxia (Cathy) Qian, Ph.D.

LinkedIn GitHub zhaoxiaq@gmail.com 267-455-3441 Seattle, WA U.S. Permanent Resident

Summary: Data-driven PhD specializing in implementing end-to-end machine learning solutions in production environment. Looking for opportunities to leverage **analytics, machine learning and software engineering** expertise to make impacts.

KEY SKILLS

Programming: Python (Pandas, NumPy, scikit-learn, Tensorflow, PyTorch, NLTK, OpenCV), SQL, PySpark, C#
Machine Learning: search ranking, demand forecasting, recommender system, deep learning, natural language processing
Data Mining: probability and statistics, quantitative analysis, A/B testing, Snowflake, Tableau, Omniture
Software Engineering: Docker, Airflow, Kubernetes, Kubeflow, Jenkens, Qubole, Apache Spark, AWS, Azure, Git
Specialties: project management, product sense, cross-departmental collaboration, interdisciplinary innovation

PROFESSIONAL EXPERIENCE

Machine Learning Scientist II, Expedia Group, Seattle, WA April 2019 – present

- **Demand Forecasting:** Led the first end-to-end machine learning solution to forecast booking id level car pickup rate.
  - Built business case, carried out data mining and analysis of 1 million entries, extracted and synthesized about *100* features and selected *45* based on feature importance, performed hyperparameter tuning of tree-based models.
  - Delivered an optimized XGBoost model with *a logloss of 0.58* and *ROC-AUC score of 0.75*, which decreased revenue prediction error from ***18.3% to 2.7%,*** resulting in ***$2.9 million*** improvement in annualized net revenue.
  - Tested and productionized code with version control and deployed the pretrained model as an API hosted on AWS.
  - Actively monitored data and model drift in production environment with a dashboard; optimized model performance by retraining and updating models regularly without additional engineering support.
- **Search Ranking:** Implemented a learning to rank algorithm to sort cars in search results; performed exploratory data analysis to identify and synthesize important features; developed a backward feature selection method that has quickly eliminated *60%* of features without degrading model performance; achieved estimated annualized revenue increase of ***$1.4 million*****,** which is ***5x*** that of the gain from the baseline model.
- **Product Management:** Defined quarterly strategy and roadmaps for the data science team, wrote business whitepapers and technical documentations, collaborated with engineering, product and marketing teams seamlessly to guarantee successful shipment of machine learning products.

Machine Learning Engineer, Terawe Corporation, Bellevue, WA July 2018 – Mar 2019

- Designed, built and tested frontend and backend of applications for users to get medical image classification and sentiment analysis results instantaneously using Python, Flask, SQLite, SQLAlchemy, HTML and CSS; packaged apps into Docker image containers and pushed them to DockerHub to allow for team accessibility.
- Implemented intelligent conversational bots using Microsoft Bot Framework in C#; authored samples and contents of Microsoft Azure service for developer audience with a Microsoft team.
- Built and optimized numerous predictive machine learning models to classify images, texts, time series and audio data in Python on Microsoft Azure Machine Learning platform; served as an internal machine learning consultant.
  - **Image Classification:** Built a ResNet50 neural network with customized fully connected layer to identify cervical image type; experimented with image cropping, image transformation, data augmentation, dropout and customized CNN and VGG16 models; achieved an optimized classification accuracy of 70%.
  - **Sentiment Analysis**: Crawled tweets using Twitter API and implemented an airline Twitter sentiment analysis model using natural language processing techniques and logistic regression, achieved classification accuracy of *0.77* and f1 score of *0.76*; built a sentiment analysis app with input as targeted airline and output as sentiment classification.

Data Science Fellow, Springboard, Remote (GitHub) June 2017 – July 2018

- Wrangled and extracted information from millions of records from databases, XML and JSON files using SQL and Python.
- Performed hypothesis test, calculated confidence intervals and statistical significance using SciPy statistical methods.
- Constructed features from images, text, categorical and numerical data via mapping, clustering, and factorization; optimized logistic regression, gradient boosting and deep learning models for classification and regression problems.

Washington Research Foundation Innovation Fellow, University of Washington, Seattle, WA April 2015 – May 2017

- Iteratively formulated simulation-driven hypothesis using simulation-experiment-learning cycles; tested hypotheses by performing experiments and feeding experimental results back into simulations.
- Lead a group of five researchers and published research results in three high impact journals. (Google Scholar)

EDUCATION

Ph.D., Physical Chemistry, University of Pennsylvania, Philadelphia 2014
B.S., Physical Chemistry, University of Science & Technology of China, Hefei 2009
Computer Science Certificates**:** Algorithms: Design and Analysis, Machine Learning, Deep Learning, Applied Text Mining