Jingxian Lin

Tel: 951-275-1055
Location: San Jose, CA
Medium: medium.com/@jingxianlin
Email: 2020jxlin@gmail.com
GitHub: github.com/JingxianLin
LinkedIn: linkedin.com/in/jingxian-lin

Status

Permanent resident

Education

- **PhD in Physics**, University of California, Riverside, California, 2006
- **MA in Statistics**, University of Michigan, Ann Arbor, Michigan, 2008
- **Self-Driving Car Engineer Nanodegree**, Udacity, 2017
- **Data Scientist Nanodegree & Machine Learning Engineer Nanodegree,** Udacity, 2020

Skills

- **Programming:** Python (sklearn, pandas, numpy), R, Java, SQL, SPSS, Matlab, Spark/MapReduce
- **Machine Learning:** Classical & Penalized Regression Methods (Lasso, Ridge), Decision Tree, Random Forest, Gradient Boosting Machine, Regularization, Clustering, K Nearest Neighbors, K-means, Multi-class Adaboost, PCA, Feature Engineering, Computer Vision and Deep Learning
- **Tools:** MySQL, Spark, Hadoop, TensorFlow, Keras, PyTorch, AWS, Tableau, etc.

Work Experience

Data Analyst, China Mobile Technology (USA) Inc., Milpitas, CA **2014 - present**

- Designed Ensemble Learning algorithm for customer churn prediction & deployed model as service
- Established user generated content recommendation engine for China Mobile “MiGu” Entertainment
- Built Text Convolutional Neural Network (TextCNN) model for short text message classification and reached 96.7% accuracy on test set
- Constructed a predictive Multiple Layer Perceptron model to detect network intrusion

Data Analyst, Glogou, Inc., Santa Clara, CA **2012 - 2013**

- Implemented sales signal processing web services for international marketing
- Developed a sales channel recommendation system using collective intelligence programming

Research Assistant, Statistics Department, University of Michigan, Ann Arbor, MI **2007 - 2011**

- Applied variable selection ideas to compute principal component functions in interpretable ways
- Proposed a data-driven method to select fixed and random effects in linear additive mixed model

Projects

Customer Churn Prediction in Telecommunications Industry

- Developed algorithms for telecommunications service vendors to predict customer churn probability based on labeled data via Python programming and Apache Spark.
- Preprocessed data set by data cleaning, categorical feature transformation and standardization, etc.
- Trained supervised machine learning models including Logistic Regression, Random Forest and K-Nearest Neighbors, and applied regularization with optimal parameters to overcome overfitting.
- Evaluated model performance of classification via k-fold cross-validation technique and analyzed feature importance to identify top factors that influenced the results.

User Generated Content (UGC) Recommendation Engine Development in Apache Spark

- Built data ETL pipeline to analyze UGC rating dataset and conducted online analytical processing with Spark SQL.
- Implemented Alternative Least Square model to provide personalized UGC recommendation and developed user-based approaches to handle system cold-start problem.
- Conducted model hyper-parameters tuning with Spark ML cross-evaluation toolbox and monitored data processing performance via Spark UI on AWS.

Self-Driving Car Engineer Nanodegree Capstone Project

- Used Convolutional Neural Networks to classify traffic sign and predict steering angle from images.
- Applied Computer Vision and Machine Learning techniques for vehicle detection and run the final project on an actual self-driving car (Ford Lincoln MKZ).