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

Education

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

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, A/B Testing, etc.

Work Experience

Data Analyst, China Mobile Technology (USA) Inc., Milpitas, CA   **2014 - present**
- Designed Ensemble Learning algorithms for customer churn prediction and deployed the trained models as a service to Container Service and Kubernetes Service
- Established User Generated Content (UGC) recommendation engine for China Mobile "MiGu" Entertainment in Apache Spark
- Built Text Convolutional Neural Network (TextCNN) model for short text message classification and reached 98.5% accuracy on test set
- Constructed a predictive Multiple Layer Perceptron (MLP) model to detect network intrusion
- Worked closely with the rest of the team to make sure systems are highly scalable and reliable

Data Analyst, Glogou, Inc., Santa Clara, CA   **2012 - 2013**
- Implemented sales signal processing web services for international marketing and a sales channel recommendation system using collective intelligence programming
- Developed a website of top ad agencies based on good Search Engine Optimization (SEO) techniques to attract 30% more traffic
- Leveraged big data techniques to gather large-scale business data, perform data engineering and utilize Tableau to visualize analytical result

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, standardization and feature correlation, 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 confusion matrix.
- Analyzed feature importance to identify top factors that influenced the results.
- Integrated with service platform providing suggestions on churn rate with accuracy of 86%.

User Generated Content (UGC) Recommendation Engine Development for China Mobile "MiGu" Entertainment in Apache Spark

- Built data ETL pipeline to analyze user generated content rating dataset and conducted online analytical processing (OLAP) with Spark SQL.
- Implemented Alternative Least Square model to provide personalized user generated content 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.

Natural Language Processing (NLP) and Text Convolutional Neural Network Modeling on Short Text Message Dataset

- Preprocessed text by tokenizing, stemming and stop-words removing, and extracted features by term frequency-inverse document frequency (TF-IDF) approach.
- Utilized text convolutional neural network (CNN) model to detect spam message and obtained 98.5% accuracy by tuning parameters.

Machine Learning Engineer Nanodegree Capstone Project Based on Deep Learning (TensorFlow)

- Created different neural network (NN) architectures to classify German Traffic Sign images.
- Identified ResNet for transfer learning after researching and comparing DenseNet, VGG and ResNet.
- Achieved a classification accuracy of 97% by combining transfer learning and model fusion methods.

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 ran the final project on an actual self-driving car (Ford Lincoln MKZ).