Jingxian Lin

Tel: 951-275-1055   Email: linjingxian07@gmail.com
Location: San Jose, CA   GitHub: github.com/JingxianLin
Medium: medium.com/@jingxianlin   LinkedIn: linkedin.com/in/jingxian-lin

Education

- **Data Scientist Nanodegree & Machine Learning Engineer Nanodegree,** Udacity
- **Self-Driving Car Engineer Nanodegree**, Udacity
- **MA in Statistics**, University of Michigan, Ann Arbor, Michigan   09/2007-08/2008
- **PhD in Physics**, University of California, Riverside, California   09/2000-03/2006
- **BS in Physics**, University of Science and Technology of China   09/1993-06/1997

Skills

- **Programming:** Python, R, Java, C++, SQL, Matlab, Linux, Windows, AWS, Azure, Spark, MapReduce, Hadoop
- **Statistics, Machine Learning:** Statistical modeling and testing, ETL, Clustering, Classification, Regression, Random Forest, KNN, AdaBoost, PCA, Reinforcement Learning, Deep Learning, Neural Networks, Computer Vision, CNN, RNN, NLP, XGBoost, GANs
- **Tools:** MySQL, MongoDB, Hive, TensorFlow, Keras, PyTorch, Scikit-learn, Numpy, Jupyter, OpenCV, Tableau, A/B Testing

Work Experience

Data Engineer, Meta Platforms Inc., Menlo Park, CA   08/2021 – 03/2022
- Launched new data models and ETL processes to estimate item-level and labeler-level performances
- Delivered the best data foundation that drives impact through informed decision making
- Setup Feature Stores by minimizing duplication of data engineering efforts, accelerating the machine learning lifecycle, and opening up new levels of collaboration across data science teams

Data Analyst / Research Engineer, China Mobile Technology (USA) Inc., Milpitas, CA
12/2014 – 10/2020
- 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
- Enhanced efficiency of a face recognition system by integrating a KNN based model
- Developed a vehicle detection system by utilizing a Histogram of Oriented Gradients (HOG) feature extractor and a Linear SVM classifier
- 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

Research Assistant, Statistics Department, University of Michigan, Ann Arbor, MI   09/2007 – 08/2012

- 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
- Analyzed endpoints for plasma etch applications based on statistical models and provided statistical consulting to scientists

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.