

# Guangqiang Lu

Dalian China | 17604264740 | gqianglu@outlook.com

## EXPERTISE

- Data and Quantitative Analysis
- Model Analytics
- Predictive Modeling
- Data-Driven Prediction
- Big Data Queries
- Data Mining and Visualization Tools
- Machine Learning Algorithms
- Research and Reports

## EXPERIENCE

**Accenture Dalian**(Third party solution provided company)

**Data Engineer:** 2018.10 to Present NewIT Big Data departure

- MultiOutput Regression model: Use multioutput RFR to predict next 7 days volume based on Error data, **RMSE: 13.01**(mean: 12.90)

**Weishenma Tec Dalian**(Providing Tech to small bank for Risk Contral)

**Algorithm Engineer:** 2017.09 to 2018.09 Big Data departure

I use Spark, TensorFlow, Keras, scikit-learn, etc. libraries to complete machine learning and deep learning algorithms analysis and modeling. From feature engineering, model building, model evaluation, model analysis and model application independently.

- User credit model: Use collaborative filtering ALS algorithm to recommend the completion of the user implicit variables, the score of 51 classes for **Multi-Label** using Stacked LSTM algorithm to achieve accuracy: **98.87%**.
- Taobao order classification model: Apply NLP algorithms for data processing such as TF-IDF and Word2Vec, and use LSTM, CNN, ResNet, Wide & Deep algorithms modeling and achieve parametric comparison, finally use ResNet to achieve accuracy: **92.20%**.
- UnionPay-Yidun users scoring model: Spark Streaming **real-time** scoring.
- Personal credit scoring model: Developing the SMOTE algorithm independently based on the problem of data volume imbalance and the sample generation algorithm based on the negative sample, AUC is increased by **31.15%** and KS is increased by **65.71%**.
- Asset-end credit scoring model: Using XGBoost and Spark to build the model, the model KS value: **0.44**.

**Neusoft Dalian**(Third party solution provided company)

**Algorithm Engineer:** 2016.07 to 2017.08 Big Data departure

I use Spark and Hadoop ecosystems related technologies to build big data machine learning models.

- Dalian University of Technology teachers clustering model: Use KMeans to cluster the population feature variables, the clustering accuracy: **93.00%**;
- Liaoning Unicom user gender classification model: Use logistic regression+PCA, random forest and naive Bayes algorithm modeling, tuning and cross-validation, accuracy: **73.50%**;
- Jiangxi Unicom monthly income forecast model: Use BP neural network for Jiangxi Unicom monthly income forecast, accuracy: **97.57%**;
- Monitoring video traffic statistics model: Use CNN and combined with SVM for video recruitment traffic statistics, the accuracy: **85.10%**;
- Using TensorFlow to implement SSD and YOLO algorithms for real-time object detection, using Caffe and TensorFlow for computer object detection.

**Zhongkeruan Tech Dalian**(Third party solution provided company)

**Java Engineer:** 2015.11 to 2016.06 Bank and Insurance departure

Use Java to develop and test Bainian Life Core Business System, add and modify business functions on core system, skillfully apply SQL for Oracle database DML, optimize SQL and multi-threaded application.

## EDUCATION

Northeastern University | Bachelor Degree in Mathematics

## SKILLS

**Machine Learning Framework:** TensorFlow, SparkMLlib, Keras, scikit-learn, Caffe, PySpark, MXNet

**Big Data Processing:** SparkStreaming, SparkSQL, Hive, HBase, Impala, Hadoop

**Languages:** Python, Scala and Java

**Algorithms:** Logistic Regression, SVM, Decision Tree, Random Forest, Naïve Bayes, KMeans, CNN, LSTM

**Competitions:** Participated in **Kaggle Machine Learning** Contest **3 times** ( Top 6%, Top 9% and Top 14%)

**Others:** GitHub open source project **Scikit-Plot** contributors(link: <https://github.com/reiinakano/scikit-plot>)