Guangqiang Lu

Dalian China | 17604264740 | gqianglu@outlook.com

Work Experience

Accenture Dalian

Data Scientist: 2018.10 to Present   NewIT Big Data department

- **Automatic Machine Learning Framework:** Based on Scikit-learn and TensorFlow, implemented **automatic** data loading, feature engineering, model building, model training, model selection, model evaluation, and model RESTful API calls as **Auto-ML** framework **independently**. Supports Google Cloud GCS, local files even with in-memory objects as training and testing data, **3 lines of code** can be used for all Pipeline implementations, open-sourced to PyPI (auto-ml-cl), can be installed within **different platforms** by using pip.
- **Accenture HR Recommendation System:** Used NLP to process Accenture projects' and candidates' data, trained TF-IDF, LSI models to build a **recommendation model** to recommend projects also with candidates for the Accenture HR system. Reduced training time from **5 hours to 1 hour** by using Spark, now is already available in North America.
- **Multi-Output Regression Model**: Used multi-output Random Forest Regressor to predict next 7 days sales volume based on Ferror data, **RMSE: 13.01**(mean: 12.90).

Weishenma Tech Dalian

Algorithm Engineer: 2017.09 to 2018.08   Big Data department

I use Spark, TensorFlow, Keras, scikit-learn, etc. to implement machine learning and deep learning algorithms and model training. From feature engineering, model training, model evaluation, model analysis and model deployment **independently**.

- **User Credit Model**: Used collaborative filtering ALS algorithm to recommend items based on user implicit information, the score of 13 classes for **Multi-Label** using Stacked LSTM algorithm to build model, achieved accuracy**: 98.87%**.
- **Taobao Order Classification Model**: Applied NLP algorithms for data processing such as TF-IDF and Word2Vec, by using LSTM, CNN, ResNet, Wide & Deep algorithms to build models combined with parameters tuning, finally using ResNet achieved accuracy**: 92.20%**.
- **UnionPay-Yidun Users Scoring Model**: **Real-time** scoring based on UnionPay-Yidun data using Spark Streaming.
- **Personal Credit Scoring Model**: Developed SMOTE algorithm independently for imbalanced data problem, came up with new sample generation algorithm based on negative samples, AUC is increased by **31.15%**, KS is increased by **65.71%**.
- **Asset-End Credit Scoring Model**: Use XGBoost and PySpark to build the model, model KS value: **0.44**.

Neusoft Dalian

Algorithm Engineer: 2016.07 to 2017.02   Big Data department

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

- **Dalian University of Technology Teachers Clustering Model**: Used K-Means to cluster users' online web's data and mobile texts' data as feature variables to cluster teachers, clustering accuracy: **93.00%**.
- **Liaoning Unicom User Exception Detection Model**: Used Logistic Regression + PCA, Random Forest and NB algorithms to build and tunning models with cross-validation, accuracy: **73.50%**.
- **Jiangxi Unicom Monthly Income Forecast Model**: Used BP neural network based Jiangxi Unicom monthly income data to forecast month volumn, RMSE: **1.52.**
- **Monitoring Video Traffic Statistics Model**: Used CNN and combined with SVM for video traffic head-counts statistics, accuracy: **85.10%**.

Education

Northeastern University | Bachelor Degree in Mathematics

Technical Skills

Machine Learning Framework: TensorFlow, Spark MLlib, Keras, Scikit-learn, PyTorch, MXNet, Gensim, Spacy

Big Data Processing: PySpark, Spark SQL, Hive, Spark Streaming, Redis, Google Cloud Platform

Languages: Python, Scala and Java

Algorithms: LR, SVM, Decision Tree, Random Forest, GBDT, Naïve Bayes, K-Means, DNN, CNN, LSTM etc.

Certification: Azure Data Scientist Associate

Azure AI Engineer Associate

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

Auto-ml-cl open source project **owner** (https://github.com/lugq1990/auto-ml-cl)