**­­**

**Wei Liu**

https://github.com/wliu40 | 315-560-8306 | wliu0828@gmail.com

**Summary of Skills**

• Proficient in Python/pyspark, R, C++, Java, MATLAB, Linux.

• Complete knowledge of statistics, data structures, algorithms and object-oriented programming.

• Working/academic experience with various machine learning algorithm/problems, include regression, classification, clustering, abnormal detection, timeseries prediction, NLP and deep learning (DNN, CNN, RNN).

• Proficient using python/R libraries: scikit-learn, numpy, pandas, pyspark, matplotlib, statsmodel,NLTK, tensorflow

• Proficient with data cleansing and engineering technique using python libraries and spark.

• Familiar with SQL/NoSQL (MySQL, MongoDB).

• Hands on experience on AWS machine learning.

• Familiar with Big Data technologies, working with SparkSQL for data retrieval.

• Scrum/Agile development methodology certification.

• Passion of learning new technologies in ML/DS/AI area, MOOC enthusiast.

**EDUCATION**

Syracuse University, College of Engineering & Computer Science 2015-2017

*Master of Science in Computer Science* GPA:3.73/4.00

Syracuse University, College of Engineering & Computer Science 2012-2016

*PhD in Mechanical & Aerospace Engineering* GPA:3.73/4.00

Nano Degree in Deep Learning (Udacity) 2017

Deep Learning Specialization (Coursera) 2018

**Work Experience**

**Data Scientist** 2018.1-present

Discuss and meeting with business partners for understanding the problem and their expectation;

Communicate with data engineers, and data retrieval from Apache Hive with Spark SQL;

Data exploration/preparation/preprocessing/visualization, include cleansing, feature engineering, data augmentation, and etc., using spark/pandas/R/numpy, working with big-date with billions of records;

Machine learning algorithm design/exploration/comparison with spark/scikit-learn/statsmodels/tensorflow/R;

Machine learning model training, model evaluation using various metrics, present and meet with BP about results;

Ad-hoc analysis, code peer review, QA and productionization;

Working on multi-projects at the same time, including rail-wear predictions, locomotives anomaly detections, radio transmitter health predictions, collaborate with colleges for different modules;

*Norfork Southern Corp.*

**AcaDEMIC Projects**

**Deep Learning with CNN & RNN**

•Constructed convolutional neural network with 60,000 images of 10 different objects from CIFAR-10 as training data

•Built LSTM RNN models with Tensorflow and generated Simpson TV scripts with existing scripts as training dataset.

•Accelerated the training process using GPUs by deploying the model in the FloydHub and Amazon EC2 instance.

**Recommendation Engine**

•Built movie recommendation engine with: 1) memory based collaborative filtering algorithms, 2. Latent factor matrix factorization model with SGD optimizer & L2 regularization, 3. ALS matrix factorization method in Spark.

**Amazon AWS machine learning and Cloud computing**

•Trained Linear Regression and Multiple Logistic Regression models with default & customized recipe. Performed ML model quality evaluation with confusion matrix and interactive cutoff threshold adjustment.

**Natural Language Processing**

•Text classification (20newsgroups) with sklearn(NB, SVM) and Spark (NB). The highest accuracy of 97.6% was achieved by multinominal NB model in Spark.

•Built a text generator with LSTM-RNNmodel which was trained by a popular Chinese novel.

•Built language translation model (English to French) using sequence to sequence RNN in Tensorflow.