**­­**

**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.

• Proficient 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**

**Degree:**

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

Chongqing University (China), ***Bachelor*** *in Physics 2009*

**Certifications:**

Nano Degree in Deep Learning (Udacity)

Deep Learning Specialization (Coursera)

**Work Experience**

**Data Scientist @** *Norfork Southern Corp.* 2018.1-present

Discuss and meeting with business partners for understanding the problem and their expectation, work with data engineers, retrieving data from Apache Hive with Spark SQL;

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

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

Machine learning model training, hyper-parameter tuning, model evaluation/improvement using various metrics, Ad-hoc analysis, code peer-review, QA and productionization; Working on multi-projects concurrently, including rail-wear predictions, locomotives anomaly detections, radio transmitter health predictions, collaborate with colleagues for different modules via git version control;

**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 models with default & customized recipe. Performed ML model quality evaluation with confusion matrix, interactive cutoff threshold adjustment and other hyper-parameters.

**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 popular Chinese novels.

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