**­­**

**Wei Liu**

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

**Summary of Skills**

• Proficient in Python/pyspark, Familiar with R, C++, Linux.

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

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

• Proficient with Big Data technologies (Spark, Hive, HDFS, MapReduce)

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

• Hands on experience on AWS machine learning.

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

**Work Experience**

**Data Scientist @ Norfork Southern Railways Corp.** Jan.2018-present

**Big-data analytics with Spark:**

•Anomaly detection of VSWR/Losing cooling water/Losing Lubricant oil of locomotives/ Rail-wear prediction

Retrieving data from HDFS using Hive/Spark, applied/design different ML algorithms, include SVM, Linear Regression,

K-means, Time Series decomposition/forecasting, Anomaly detection, RNN, Large Scale Geographical matching, Change-point

Detection and etc. with python and Spark.

**Computer vision and deep learning:**

•Intermodal container OCR (natural scene image) recognition, Railway tie-plate and spikes recognition.

Apply and developed deep-learning algorithms for image semantic segmentation using Opencv, FCN (Transfer learning),

Yolo and etc. with Tensorflow and keras.

**Tool and libraries:**

•Hive, spark, numpy, pandas, stats models, scikit-learn, tensorflow, keras, opencv, basemap, folium, matplotlib and etc.

**Other job responsibilities:**

•Project feasibility research, Data exploration/cleansing, Ad-hoc analysis with big data, peer code-review, QA analysis,

Apache NIFI flow design and code productionization.

**EDUCATION**

**Degree:**

Syracuse University, College of Engineering & Computer Science 2017

***Master of Science*** *in Computer Science*

***PhD*** *in Mechanical & Aerospace Engineering*

Chongqing University 2009

***Bachelor*** *in Thermo-physics* Engineering

**Certifications:**

Nano Degree in Deep Learning (Udacity)

Deep Learning Specialization (Coursera)

**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.

**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.