**­­**

**Wei Liu**

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

**Summary of Skills**

• Proficient in C++, Python, Java, MATLAB, JavaScript, HTML, CSS, XML, R, Linux.

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

• Hands on experience and specialized training with machine learning algorithm, include regression, classification, clustering, abnormal detection, collaborative filtering, NLP and deep learning (DNN, CNN, RNN, GAN).

• Familiar with scikit-learn, numpy, pandas, pyspark, Apache Spark 2.x, NLTK, Tensorflow, xgboost

• Familiar with data scraping, data processing, data visualization with Python and R (Shiny).

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

• Hands on experience on AWS machine learning, S3, EC2, and Google Cloud.

• Knowledge with Big Data technologies (Hadoop, Map Reduce, Pig, Hive, Hbase, Cassandra).

• Scrum/Agile development methodology certification.

**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

Data retrieval, cleansing and engineering, machine learning model training, ad-hoc analysis and results presenting

*Norfork Southern Corp.*

**Teaching Assistant (Mathematical Analysis, Auto CAD, etc)** 2012-2016

*Syracuse University, College of Engineering & Computer Science*

**AProjects**

**Deep Learning with CNN & RNN**

•Constructed convolutional neural network with 60,000 images of 10 different objects from CIFAR-10 as training data, the highest accuracy of 62.2% was achieved by hyperparameter tuning.

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

•Faces image generation using deep convolutional generative adversarial networks on 200,000 images of training data.

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

**Recommendation Engine**

•Built movie recommendation engine with different methods: 1) memory based collaborative filtering algorithms (RMSE = 1.01). 2. Latent factor matrix factorization model with SGD optimizer & L2 regularization (RMSE = 0.98). 3. ALS matrix factorization method in Spark (RMSE = 0.95).

**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.

**Flight Company Database Design and Implementation**

•Simulated the flight company database from scratch and design the dimensional model. 20+ practical functional views, functions and procedures were created with MySQL.