**Zihao Wang**

**wqm1800@gmail.com | (551)225-9955 | 204 10th St, Apt. 314, NJ 07302 | https://zw1074.github.io**

**Educations**

**New York University,** Courant Institute of Mathematics **New York, US**

*Master of Science in* ***Data Science (3.80/4.0)*** *May 2017*

**Nanjing University Nanjing, China**

*Bachelor of Science in* ***Computational Mathematics (3.50/4.0)*** *Jun 2015*

**Related Courses**

Machine Learning, Probability Theory, Stochastic Processes, Statistical Inference, Business Understanding for Data Science, Big Data, Optimization Theory, Deep Learning, NLP, Advanced Python, Fundamental Algorithm, Artificial Intelligence, Operating System, Linear Algebra.

**Skills**

* Programming Skills: Python (Proficient), C/C++ (Proficient), SQL, Hadoop, Spark, AWS, Git, Matlab
* Data Analytic Skills: Machine Learning, NLP, Deep Learning, Web Scrawling, Data Visualization
* Packages: scikit-learn, genism, nltk, Theano, TensorFlow
* Languages: English, Chinese

**Professional Experience**

**Data Scientist New York, US**

*National Grid Sep 2017-Present*

* Design and implement a prototype of AI Chatbot by AWS Alexa, Lex and Lambda functions.
* Design a location intelligent package that can help to regularize street name, geospatial information conversion and analyze geospatial data.
* Help to build an end-to-end system for auto-renewing NYC permits based on work order and paving order.
* Build a system that can efficiently connect polygons with poly-lines by building a KD-tree based on geospatial information.

**Research Assistant, Intern New York, US**

*American International Group Inc. Aug 2016-Dec 2016*

* Contribute to the license plate detection on low-resolution image dataset and heat map generation of damaged parts.
* Build an efficient license plate detector by fine-tuning a simple but efficient convolution neural network to generate saliency map and then use OpenCV to extract contour of license plate.
* Build an end-to-end Grad-CAM solution to extract information about damage part from a pre-trained model and an image.
* The license plate detector achieves 15% top-5 error with 30s/image speed.

**Academic Projects**

**Center for Data Science, NYU New York, US**

*Machine Learning: Duplication Detection for health care information system Feb 2016-May 2016*

* Build an end-to-end solution to detect duplicated record in health care information system.
* Find sets of similar information entries by using K-means clustering and use elbow method to determine the number of sets.
* Create our own parallel filter algorithm to find possible duplication pairs in each set.
* Overall performance is 95% accuracy.

*Machine Learning: Yelp Restaurant Rating Prediction Sep 2015-Dec 2015*

* Build a model to predict future rating level of restaurant based on business attributes, previous ratings and Yelp reviews.
* Divide the rating into 3 different levels based on the distribution to make a balanced dataset.
* Use Google pre-trained word2vec model to represent Yelp reviews as 300-dimension vector.
* The best model is a logistic regression with L1 regularization and its average cross-validation AUC-scores is 0.86.

*Anomaly Detection: Real Time Data Ingestion and Anomaly Detection for Particle Physics* *Sep 2016-Dec 2016*

* Build an anomaly detector for detecting abnormal events from 100GB normal CERN particle collision event.
* Train a compressor and re-constructor for capturing insight feature of physics events by using deep neural auto-encoder.
* Evaluate auto-encoder by calculating reconstruction error (R2-score) and deep MLP auto-encoder has 0.95 R2-score.
* Detect abnormal event by checking whether reconstruction error from well-trained auto-encoder is greater than threshold.
* This work is supervised by Prof. Kyle Cranmer and presented in NIPS 2016 invited talk.