Lizhen Tan

lt911@nyu.edu | 646.824.6362 | New York, NY

EDUCATION

NEW YORK UNIVERSITY
MS in Data Science
Expected May 2017 | New York, NY
Cum. GPA: 3.75

STONY BROOK UNIVERSITY
BS in Astronomy/Planetary Science
& Applied Mathematics And Statistics
Expected May 2013 | Stony Brook, NY
Cum. GPA: 3.91

LINKS

LinkedIn:// lizhentan
Webpage: Liz-page

COURSEWORK

Deep Learning
Machine Learning
Natural Language Processing
Time Series
Causal Inference
Big Data

SKILLS

COMPUTER SKILLS/TOOLS

Linux • Python • R
PyTorch • LaTeX • GitHub
Amazon AWS • MySQL • TensorFlow
Hadoop • Microsoft Office
Spark (beginner)

LANGUAGE

Fluent:
English • Chinese (Cantonese and Mandarin)

LEADERSHIP

Graduate:
Fall 2016 - Spring 2017
Secretary in Leadership Circle of CDS (Center for Data Science)

Spring 2017
Grader of Python class

Undergraduate:
All semesters
Dean's List
Fall 2012 - Spring 2013
Research intern (Astronomy)

WORK EXPERIENCE

CIVITAS LEARNING | Associate Data Scientist (Internship)
Summer 2016 | Austin, TX

- Collaborated with another intern on a new project for building predictive models on students' success based on their engagement in Learning Management Systems. (e.g. Blackboard, Oracle, etc.)
- Acquired data from data warehouse by performing SQL queries
- Analyzed top features of models for insights of students' engagement pattern

ACADEMIC PROJECTS

BUILDING IMAGE GENERATOR USING CONDITIONAL DCGAN Spring 2017 | New York University

- Controlled image randomness by adding class information to DCGAN model.
- Processed class context information into either one-hot embedding or pre-trained GloVe word embedding.
- Generated images were reasonable by feeding both MNIST and CiFar10 datasets. (better results using one-hot embedding)

SEMI-SUPERVISED LEARNING ON MNIST HANDWRITTEN DIGIT RECOGNITION Spring 2017 | New York University

- Utilized CNN(convolutional neural network) architecture to get image features
- Increased the small labeled dataset size using data augmentation (e.g. image scaling, image rotating, etc.)
- Improved model by applying a psudo-label semi-supervised learning approach to the limited labeled data. (Final accuracy of this framework had a 1% boosting)

PREDICT AND QUANTIFY DEVELOPMENT INFLUENCE ON REAL ESTATE VALUES Fall 2016 | New York University

- Quantified influence of new transactions of commercial units on a pre-defined neighboring real estate values
- Applied machine learning techniques (Logistic Regression, Random Forest) to further explore the potential building features which had the most impact in the target (whether an influence existed)

UNDERSTANDING AND LEARNING AN AUTOMATED QUESTION ANSWERING SYSTEM Spring 2016 | New York University

- Applied text analysis on questions asked on Yahoo!Answers : explored several multi-class classification methods to predict the question category (4 categories were used in the project)
- Built a program to output an answer to a newly asked question, where the answer was found from a similar archived question in the predicted category

ANALYSIS OF INTERACTION AMONG DIFFERENT TAXI MODES IN NYC Spring 2016 | New York University

- Conducted data cleaning and data extraction using Map-Reduce on Hadoop
- Analyzed the interaction between newly introduced taxi modes (Uber, green taxis) and yellow taxis

TWITTER SENTIMENT ANALYSIS Fall 2015 | New York University

- Acquired data by scraping TV show related data from Wikipedia
- Performed sentiment analysis on Twitter Tweets (dataset: Sentiment140)
- Built predictive model to examine impact of popularity in social media (derived feature from sentiment analysis) on TV show renewal