

# YIDI ZHANG

16 Stuyvesant Oval #4G, NY 10009 | (646) 209-4733 | yz3464@nyu.edu

## EDUCATION

<b>New York University</b> , New York, NY M.S. in Data Science (Introduction to Data Science, Programing for Data Science, Statistical and Mathematical methods, Machine Learning, Big Data, Deep Learning ( <b>Yann LeCun</b> ) )	05/2018
<b>Shandong University</b> , Jinan, Shandong, China B.S. in Software Engineering B.S. in Accounting	06/2016 GPA: 3.83/4.0 (Major)

## TECHNICAL SKILLS

**Programming Languages:** Python, Java, C/C++, C#, HTML, CSS, Javascript, R.  
**Operating Systems:** Linux, Windows, OX.  
**Software:** Jupyter Notebook, **Hadoop**, **MapReduce**, **Spark**, **TensorFlow**, **PyTorch**, Matlab, Microsoft Visual Studio, Eclipse, Latex, Python Notebook, Oracle Database, **MySQL**, **SQL Server**, **Excel**, Microsoft Office series.

## WORK EXPERIENCE

<b>Machine Learning Intern</b> , Centiment.io, New York <ul style="list-style-type: none"><li>Worked on machine learning and deep learning algorithm to evaluate the viability of films' content over 27 data sources including IMDB, rotten tomatoes, etc.</li><li>Used advanced machine learning such as random forest tree and the support vector machine to organize, interpret, and use this data in the forecasting of how content will perform.</li></ul>	02/2017 - Present
<b>Data Analyst Intern</b> , ChangeThePresent.org, New York <ul style="list-style-type: none"><li>Gathered data from the organization's web content and used quantitative and qualitative research methods to analyze data.</li><li>Researched original web source code to resolve a logo presentation problem.</li></ul>	12/2016 - 02/2017
<b>Software Engineer Summer Intern</b> , Beijing Runtime Corp., Ltd, Beijing, China <ul style="list-style-type: none"><li>Coded automatic tests in Java and applied them to test the telecommunication database system to ensure system's reliability and stability.</li><li>Developed a web scraping application for information extraction from social media sites and normalized scrapped unstructured data into structure database to help project managers make better business decisions.</li></ul>	08/2015-10/2015
<b>Finance Intern</b> , Jin Shang Commercial Bank, Taiyuan, China <ul style="list-style-type: none"><li>Wrote financial reports on 20 companies and summarized their marketing activities for daily briefings to help leaders make strategic decisions on their investment.</li></ul>	07/2014-09/2014

## RESEARCH AND PROJECTS

<b>Image Captioning</b> , New York University, New York <ul style="list-style-type: none"><li>Built an attention based model that automatically learns to describe the content of images on MSCOCO.</li><li>Used a convolutional neural network to extract a set of features vectors referring to as annotation vectors.</li><li>Implemented LSTM to produce captions by generating a word at every time step.</li><li>Compared the performance of hard attention and soft attention based on BLEU.</li></ul>	01/2017-present
<b>Yelp Data Analysis</b> , New York University, New York <ul style="list-style-type: none"><li>Extracted useful information from yelp's reviews using scikit-learn package; performed feature selection on yelp data sets with business sense by comparing mutual information.</li><li>Used GridSearchCV to get the best configuration for selected models; Compared the AUC of models to select the best model.</li><li>Performed model evaluation using AUC.</li><li>Improve 6% on accuracy in predicting which reviews are more likely to be thought useful.</li></ul>	10/2016-12/2016
<b>NYC Vehicle Collision Analysis</b> , New York University, New York <ul style="list-style-type: none"><li>Coded a program visualizing essential information of vehicle collision in NYC over the last year and explored the weather features in conjunction with vehicle collision data.</li></ul>	10/2016-12/2016
<b>Graduation Project</b> , Shandong University, Jinan, China <ul style="list-style-type: none"><li>Coded adaptive image enhancement algorithm (C++) and enhanced inputed fingerprint images quality.</li><li>Optimized GLCM (Gray Level Co-occurrence Matrix) and led to a 25% time's saving in segmenting fingerprint images.</li></ul>	12/2015-06/2016