

## Job Objective

Data Analytics Master Student looking for a position of Data Scientist/Data Analyst, where I can apply my coding skills, including Machine Learning and Natural Language Processing to solve real word problem.

## Education

### The George Washington University

M.S. IN DATA ANALYTICS

- Current GPA: 3.83/4.00
- Related Courses: Data Analysis for Scientists and Engineers, Design and Analysis of Algorithms, **Machine Learning**, Programming for Analytics, **Database Management**, Intro to Statistical NLP

Washington, D.C., USA

Jan. 2018 - Dec. 2019

### Central South University

BACHELOR OF CIVIL ENGINEERING

Changsha, Hunan, China

Sep. 2012 - Jun. 2016

## Skills

### Software and programming Languages

**Python** (PyTorch, fastai, gensim), MySQL, **MongoDB**, **PostgreSQL**, Java, R, Linux Shell, Minitab, LaTeX, React, Docker, Git

### Machine Learning

Classification (AdaBoost, Random Forest, SVM), Regression, Clustering, Feature Engineering (Featuretools)

### Data Visualization

D3.js, hvPlot(Interactive Graphing), Semiotic(React Visualization Framework), Matplotlib, Seaborn

### Statistical Methods

A/B Testing, Hypothesis Testing and Confidence Intervals, Principal Component Analysis and Dimensionality Reduction

## Projects

### Modeling FTA for PDS intervention

FEATURETOOLS/POSTGRESQL

- Predict if a defendant will fail to appear so PDS could decide whether they should give away pre-paid metro card
- Web scraping data from CourtView
- Use Featuretools to implement feature engineering

DC Lab

Sep. 2019 - Now

### Just in Time Recommendation

DJANGO/POSTGRESQL/D3.JS/NEO4J

- Work with 4 colleagues to develop a company wiki system to shorten the training time for new employee
- Extract the relationship between document and topic, give recommendation base on the keywords that user searched
- Automatically update the topic modeling database as users upload files

CCNOAI

Jun. 2019 - Aug. 2019

### Deanonymizing author with a novel identifying feature

PYTHON/ADABOOST/RANDOM FOREST

- Use Project Gutenberg Dataset
- Aim at identifying the writing style for different authors
- Achieve 61.71% Macro F1 score for 5 authors

School

Nov. 2018 - Dec. 2018

### IMDB Reviews Sentiment Analysis

PYTHON/KERAS/LSTM

- Elaborate the theory behind LSTM
- Use GloVe as language model

School

Nov. 2018 - Dec. 2018

### Email Data Analysis Based on Text Clustering

PYTHON/LDA

- Text Analysis on email content
- Staff sometimes talk about vacation before bankruptcy. But job-hopping was the main topic during the last days at work .

School

Aug. 2017 - Aug. 2017