|  |
| --- |
| James Xue jx2181@columbia.edu • (973) 610 2618 • 134 W 70th St, New York, NY 10023 |

# EDUCATION

Columbia University

New York, NY

MS in Data Science, GPA: 3.7/4.0

Expected May 2019

2nd place in the IBM Columbia Call for Code Hackathon, teaching assistant in Algorithms course

Rutgers University

New Brunswick, NJ

BA in Mathematics, Biological Sciences, and Economics

2010 - 2014

Awards: Rutgers Presidential Scholarship (Full Ride), Edward L. Shustak Memorial Scholarship

# Projects

improving Advertisement click-through rate with gan

* Developed the data pipeline and Tensorflow based model for a Generative Adversarial Network (GAN) to improving web advertisement click-through rate for a marketing technology firm.
* Optimized GAN over several weeks of parameter search with 16 GPUs and 120GB of RAM on Google Cloud.
* Visualized complex results with PCA dimensionality reduction and by tracking input transformation.

Genetic disease phenotype Classification with computer vision

* Worked with researchers from Columbia Medical School to improve deep learning models (patent pending) for classifying cells with various genetic deficiencies.
* Obtained promising results with mitochondrial disease classification and worked to extend success to several other intracellular structures.

GPU-based Stochastic Gradient Descent for image classification

* Developed 4 different GPU-based SGD algorithms using PyCuda with C kernels.
* Showed that algorithms are train 8x as fast as the equivalent scikit-learn model with comparable accuracy on MNIST.

Steam recommendation engine

* Created neighborhood-based recommendation engines and applied Factorization Machine techniques to Steam video game customer engagement data.
* Enriched Steam dataset with video game ratings and descriptions datasets to extend recommendations.

Fuel Economy Prediction

* Using 2015, 2016, and 2017 vehicle specification data from the U.S. Department of Energy, trained models to determine fuel efficiency for 2018 vehicles with scikit-learn.
* Achieved 96% test accuracy with ridge regression and XGBoost on interactions features from 9 numerical features and 8 categorical features (determined through feature selection from 140 total features).

# Work Experience

Navigant Consulting

New York, NY

Data Science Summer Consultant

2018 - 2018

* Created interactive visualizations and simulations with ggplot2, Excel VBA, and Python to support expert witnesses and lawyers in Financial Disputes and Litigation cases.
* Leveraged Beautiful Soup and xlwings to scrape and aggregate large financial datasets and deposition transcripts.
* Replicated a published leveraged ETF model in Python.

J.P.Morgan Asset Management

New York, NY

Client Portfolio Management Analyst

2014 – 2015

* Reducing monthly data cleaning and reporting time by 85% and reconciling 99% of data automatically in Excel and VBA by matching bond data from internal and external sources on a CUSIP level and flagging potentially incorrect data.
* Interfaced with institutional clients and provided allocation, performance, and transaction analysis and visualization on over $68 billion in AUM.

# Skills

TechnologY

Python, numpy, pandas, xlwings,

PyCuda, OpenCL, C

scikit-learn, Keras, Tensorflow

R, ggplot2

Excel, VBA, SQL, HTML, CSS, JavaScript

Git, PySpark, Google Cloud