

# Liushiya Chen

liushiya.chen@gmail.com • 347 720 2013 • github.com/theshiya

Available for immediate start remotely or in New York

## EDUCATION

---

### Columbia University

New York, NY

M.S. Computer Science (Machine Learning), GPA: 4.0

Jan 2020 - Present

- Relevant courses: Advanced Algorithms, Bayesian Machine Learning, Causal Inference, Discrete Optimization, Introduction to Databases, Natural Language Processing

### University of Pennsylvania

Philadelphia, PA

B.A. Mathematics (with Computer Science & Statistics)

Dec 2017

- Top 22% Putnam Math Competition 2015, admitted to PhD program at Columbia University

## EXPERIENCE

---

### Columbia Business School

New York, NY

Research Associate

Jun 2020 - Present

- Build data pipeline in Python to clean, merge, and analyze job posting data to study labor market trends
- Perform large-scale keyword search and named entity recognition on job text to cluster jobs and identify employers
- Conduct code review with other RAs, integrate their code into main codebase, and maintain technical documentation

### GIC (Singapore Sovereign Wealth Fund)

New York, NY

Data Scientist

May 2019 - Dec 2019

- Constructed fundamental investment factors using financial documents (Form 4, 10-Q/K, 13-F etc.), corporate events data, and macroeconomic data for US mid-cap equities
- Extracted industry-specific factors, such as vertical integration, from quarterly and annual reports using NLP
- Used Python to prune/combine correlated variables, detrend times series, impute missing values, and handle outliers
- Forecasted investment returns using fundamental factors with constrained and regularized linear regression models
- Improved backtesting infrastructure for model evaluation, performance attribution, and results visualization
- Built Python data pipeline with Airflow from APIs (S&P Capital IQ, Bloomberg) to SQL database and dashboard

### Millennium Management (Hedge Fund)

New York, NY

Quantitative Data Analyst

Feb 2018 - May 2019

Summer Analyst

Jun 2017 - Aug 2017

- Analyzed intraday positions data of internal equity portfolio managers to attribute performance and evaluate risk
- Implemented linear regression and decision tree models in Python to forecast manager performance using technical indicators and qualitative characteristics
- Worked on research project to quantify value of sell-side research and alternative data; used sell-side interaction log, internal analyst ratings, and positions data to estimate contribution to profit and loss
- Improved linkage of disparate equity data sets using company names, RICs, and string similarity metrics
- Contributed to internal data science toolkit with efficient Python and SQL data processing scripts
- Attended financial training programs at Bloomberg and Morgan Stanley; advised and supervised by Dr. Prem Melville

## PROJECTS

---

- **Yelp Review Summarizer & Scraper** – created greedy optimization algorithm with doc2vec to extract semantically representative reviews from Yelp business pages; created an asynchronous Yelp review scraper as byproduct
- **Forecasting Cable TV Customer Acquisition** – constructed gamma-Weibull model to forecast customer acquisitions, incorporating external factors such as population growth, macro events, and seasonal promotions
- **Deep Learning for Virtual Autonomous Driving** – built convolutional neural network in PyTorch that detects in-game objects and maneuvers vehicle in SuperTuxKart racing game

## SKILLS

---

- **Programming** Python (pandas, scipy, sklearn, statsmodels, pytorch, spacy, seaborn), SQL, R, Java
- **Software & Tools** PyCharm, Git, Jupyter, Linux, Bloomberg, Google Cloud Platform, Tableau