

# Gaoyi Shi, FRM

[gs3012@columbia.edu](mailto:gs3012@columbia.edu) / (732)-319-3506 | LinkedIn: [www.linkedin.com/in/gaoyishi](http://www.linkedin.com/in/gaoyishi)

## EDUCATION

---

### Columbia University, School of Engineering and Applied Science

*MS, Data Science, GPA: 4.0/4.0*

Sep 2018 – Dec 2021

New York, NY

- Coursework: Machine Learning, Data Exploration and Visualization, Deep Learning, Personalization and Recommendation System

### Rutgers University, School of Arts and Science

*MS, Mathematical Finance, GPA: 3.91/4.0*

Sep 2015 – May 2017

New Brunswick, NJ

- Coursework: Derivatives Pricing, Stochastic Calculus, Numerical Analysis, Time Series Analysis, Quantitative Risk Management

### Zhejiang University, School of Economics

*BS, Finance, Honor graduate, GPA: 3.86/4.0*

Sep 2011 – Jun 2015

Hangzhou, CN

## SKILLS

---

- **Specialization:** Structured financial products, NLP - Textual data parsing, Machine learning applications
- **Technical Skills:** Python, R, SQL, C++, C#, EXCEL VBA, HTML, Bloomberg, Google Cloud Platform, AWS
- **Libraries and Frameworks:** NumPy, Pandas, Matplotlib, Scikit-learn, BeautifulSoup, NLTK, Gensim, PySpark

## EXPERIENCE

---

### Moody's Analytics

*Financial Engineer*

Jan 2018 – Present

New York, NY

- Leading an innovation NLP project for next generation data solution and platform - Textual data extraction and parsing for performance data and structuring information from documents; implementing template libraries in C# and Python for text parsing, data transformation and data feed to client-facing products DataViewer and SFPportal
- Working with structured and unstructured data, developed Python tools for web crawling, data comparison and data visualization
- Researching addition of RMBS new fields/features to templates, improving current fields calculation logic and providing clients with more data points, including delinquency, loan modification, credit supports and hedge instruments
- Working with ratings teams and desks to analyze deal structures, build cash flow models and simulate different default and prepayment scenarios to finalize models; improved SFW scripts for priority of payment and trigger related issues
- Worked on SingleSource Project in Agile development mode; integrated and synchronized data from different sources to establish internal single data warehouse, created points system for deal and tranche mapping; related SQL stored procedure, ETL, database infrastructure development work in AWS environment, collaborated with operations and QA managers to coordinate testing

### Thomson Reuters

*Quantitative Analyst*

Jun 2017 – Dec 2017

New York, NY

- Covered several clients' CLO portfolios, monitored daily price movements and risks; presented month-end quality assurance reports
- Developed market comparable pricing methodologies, implemented models and algorithms for reliable estimation of CLO discount margin (DM) and tranche pricing; built scenario-based simulation system for fair price determination
- Conducted thorough research into CLO market, studied tranche and underlying loans characteristics, rating matrix and macro features to determine potential factors for multi-factor models under different scenarios
- Cleaned and processed high volumes of raw data in Python; implemented back-testing and prediction system with machine learning algorithms for tranche DM and price; achieved prediction errors for tranche DM of <5% and for market price of <0.5%

## PROJECTS

---

### Movie Recommendation System

Sep 2020 – Dec 2020

- Built a movie recommendation engine for personalized top recommended movies based on 28 million ratings and movie meta data entries
- Constructed a front-to-back pipeline in Python for feature engineering, hyper parameter tuning, model training, testing and evaluation
- Implemented hybrid tree-based weighted model based on different scenarios with recommendation algorithms including CF, MF, SVD++ and content-based fastFM models

### NYC Bike Share During COVID-19 - Data Analysis and Visualization

Oct 2020 – Dec 2020

- Analyzed and compared NYC Citi Bike usage during COVID-19 and historical years with statistical analysis and data visualization
- Built a R bookdown book on GitHub, covering topics including user types, locations and route patterns, holidays and weather effects
- Embedded with web interactive plots and animations by JavaScript D3

### Text Mining for Seeking Alpha in Stock Market

Sep 2016 – Dec 2016

- Developed a prediction system in Python which fetched financial news via Yahoo Finance News API, parsed and vectorized news in Bag of Words; applied predictive machine learning approaches SVM regression and Lasso for discrete stock price prediction
- Simulated a systematic trading strategy based on alpha model; achieved Sharpe Ratio 0.89 and alpha 2.1