Andrew (Zhehan) Shi

(917)703-6627 zs1113@nyu.edu New York, NY Portfolio Github Linkedin

EDUCATION

NEW YORK UNIVERSITY
Center for Data Science GPA: 3.78/4.00
M.S. in Data Science (Expected May 2023)

- ***Coursework***: Optimization and computational linear algebra, machine learning, big data, algorithmic trading and quantitative strategies, deep learning.

NEW YORK UNIVERSITY
Courant Institute of Mathematical Sciences GPA: 3.60/4.00
B.A. in Mathematics, B.A. in Computer Science, Minor in Business Studies (May 2021)

- 2017-2018 Dean's List Honors in Liberal Studies (10%)
- ***Coursework***: Multivariable calculus, linear algebra, partial differential equations, data structure and algorithm, probability and statistics, regression analysis, object-oriented programming, managerial accounting, foundations of finance, natural language processing, number theory, operating system, parallel computing, applied internet technology.

EXPERIENCE

NYU Research
Researcher (February 2022 – May 2022) New York, NY

- Developed framework to price exotic and basket options by using deep learning techniques to solve Euler-Maruyama discretized forward-backward stochastic differential equations.
- Implemented the research ideas in Tensorflow under the guidance of Dr. Bernhard Hientzsch and Dr.Petter Kolm.

PROJECTS

NEW YORK UNIVERSITY New York, NY

Market Impact Model (Python) GitHub

- Worked with more than 100GB+ 3-month millisecond-level high-frequency NYSE trades and quotes tick data of more than 1000 tickers to calibrate the Almgren market impact model by applying nonlinear regression.
- Formulated the Almgren-Chriss optimal execution problem as a stochastic control problem and derived the Hamilton–Jacobi–Bellman equation and solved for the control and value function.

Covariance Matrix Estimation (Python) GitHub

- Performed covariance matrix estimator analysis by contrasting four covariance matrix estimators, including but not limited to optimal rotational invariant estimator.
- Processed a subset of 30GB+ millisecond-level high-frequency trades data to custom-build 5-min sliding-window covariance matrix estimators.

Big Data Analytics (Hive, Apache Spark) Medium

- Performed an analysis of Boston restaurants' cleanliness data using designed metrics from the data provided by Yelp and Boston government. Manipulated data using Hive, and performed regression analyses were done using Spark.

Game of Life Simulation (C, OpenMP) GitHub

- Developed a parallelized version of a cellular automaton, Game of Life, based on the principles laid out by George Conway.

Natural Language Processing (Python) GitHub

- Built a python program to tag part-of-speech using Hidden Markov Model by learning from Wall Street Journal Corpora and achieved more than 95% accuracy, very close to the accuracy of human taggers

Information Retrieval (Python) GitHub

- Created a system for an Ad Hoc Information Retrieval task using TF-IDF weights and cosine similarity scores.

Course Registration System (Java) GitHub

- Designed system for administrator to manage and students to select courses with OOP paradigm
- Implemented serialization/deserialization mechanism to ensure consistent state of system.

SKILLS

- ***Programming Languages*****:** Java, Python, SQL, NoSQL, C/C++, Git, JavaScript
- ***Library:*** TensorFlow, NumPy, Pandas, Scikit-learn, OpenMP, CUDA, MPI, Hadoop, Apache Spark
- ***Operating systems***: Mac OS, Windows, Linux
- ***Languages***: English (Fluent), Mandarin Chinese (Native)
- ***Others***: Jupyter Notebook, Latex
- ***Extracurricular***: NYU mathematics society, NYU Math Finance Group Team Leader