Feng Jiaqi

116th Street and Broadway
(+1) 858-568-0228 | jf3490@columbia.edu | Github: jiaqifeng777
LinkedIn: https://www.linkedin.com/in/jiaqi-feng-504053222

EDUCATION

Columbia University New York City, NY
Master of Science in Data Science Sep 2022 - Dec 2023

The University of California, San Diego San Diego
Bachelor, Cum Laude Honor Sep 2018 - Mar 2022
GPA: 3.862 / 4.0

- Major: Applied Mathematics

Courses: Numerical Analysis and Linear Algebra, Computational Statistics, Mathematical Analysis, Math Statistics, Applicable Maths and Computing, Differential Equations, Probability Theory, Vector Calculus, Calculus and Analytical Geometry, etc.

- Minor: Data Science

Courses: Data Management, Recommender System and Web Mining, Data Science in Practice, Data Analysis and Inference, Data Structure and Algorithms for Data Science, Theoretical Foundations of Data Science, Principles of Data Science, etc.

WORK EXPERIENCE

Tutor at Data Science Institute, UCSD San Diego, CA
Tutor Sep 2021 - Mar 2022

- Held office hours twice a week, explained course concepts and debugged for students. Answered questions on class forum
- Developed and beta-tested midterm and final projects. Created test cases and rubrics. Graded weekly assignments and exams. Communicated and provided feedback to instructor during staff meetings

Grader in Mathematics Department, UCSD San Diego, CA
Grader Sep 2020 - Dec 2020

- Graded Matlab assignment. Explained concepts and helped students with Matlab coding

INTERNSHIP EXPERIENCE

Financial Engineering Group, Huafu Securities Research Institute Shanghai, China
Quantitative Analyst Intern May 2022 - Aug 2022

- Collected 3000+ stock data and indices, reformatted and processed data. Accomplished data visualization and quantitative analysis work on important indexes (eg., Wind Quan A, CSI 300, etc.)
- Implemented and completed quantitative analysis models and pattern recognition algorithms (eg., double bottom patterns, bottom divergence patterns, etc.), recognized important patterns on 3000+ stocks and indices

Cloud Platform R&D Department, Nebula Link Technology Co., Ltd Beijing, China
Data Analyst Intern May 2021 - Aug 2021

- Designed and implemented algorithms using Python on vehicle-road cooperative data. Created new methods to analyze statistics of punctuality rate, frequency of turn signals of driving vehicles, overspeed recognition algorithms, etc.
- Processed and filtered thousands of vehicle-road cooperative data on Ali-Cloud, examining and dealing missing values, and maintaining online database
- Cooperated with other R&D staff to manage projects and maintain algorithms on Ali-Cloud platform, developed new solutions, fixed bugs, and enhanced functionality and efficiency of algorithms

Software implementation Department, Berheley Statistical Big Data Co., Ltd Tianjin, China
Data Analyst Intern Jan 2021 - Mar 2021

- Devised and completed data analysis and forecasting work of Shanghai Baoshan District with Python, sorted and processed data, and established prediction models (such as LSTM, SARIMA, and other time series prediction models)
- Filtered, processed, and analyzed 1000+ economic output data, optimized prediction model, and provided data visualization, analysis, and forecasts for customers

CSC Financial Co., Ltd Tianjin China
Data Management Intern Sep 2020 - Oct 2020

- Collected daily stock data of IPO announcements of Shanghai Stock Exchange and Shenzhen Stock Exchange (release prices, expected prices, issuance upper limit, etc.). Integrated data, managed online database

PROFESSIONAL EXPERIENCE

Self Projects San Diego, CA

Prediction Model of Whether Users Would Cook and Cook Time Oct 2021 - Nov 2021

- Examined data with 500,000 instances of recipe ratings and reviews, processed data, EDA, and created engineered features such as popularity and Jaccard similarity to predict whether a user would cook a recipe
- Built bag-of-words feature vectors with counts of 4,000 most common words in each review, extracted relevant features, and trained Ridge regressor to predict cook time
- Deployed grid search to find best parameter and optimized model performance by 20%

Prediction Model of Allegation Outcome for the NYPD Complaints San Diego, CA

Self Project Apr 2021 - May 2021

- Performed data cleaning, Exploratory Data Analysis, and analysis of missing values with data of 12,000 complaints published by NYPD in 2017 with Python
- Established prediction model with 5 engineered features and RandomForestClassifier to classify and predict outcome of allegations
- Selected best parameters using cross-validation, and increased accuracy rate of prediction model by 10%

Analysis of the Effect of Maternal Smoking Behavior on New Births San Diego, CA

Self Project Mar 2021 - Apr 2021

- Studied influence of maternal smoking behavior on fetal development with physical data of 1237 pregnant women, records of smoking behaviors during pregnancy, and weights of new births
- Researched impact of other physical factors of pregnant women on weight of newborns, performed statistical tests on 3 factors and accomplished data visualizations using R

Analysis of Housing Price Data in Boston and Related Influential Factors San Diego, CA

Group Project Sep 2020 - Oct 2020

- Implemented EDA and data visualization using R with 1000+ Boston housing data. Performed statistical tests on influential factors. Examined influence of factors on housing price. Built linear models to predict housing prices

Machine Learning Image Recognition Algorithm San Diego, CA

Self Project Jun 2020 - Aug 2020

- Designed machine learning algorithm to realize image recognition using Python
- Split image dataset into training and test set, studied and adopted a technique to calculate pixel differences in pictures to distinguish between human faces and cartoon faces with accuracy rates over 99%

SKILLS

Technical Skills

Programming: Python, SQL, Java, R, Matlab
Packages: Numpy, Pandas, Seaborn, Matplotlib, Tushare, Scikit-learn,
Database: SQLite, PostgreSQL, MySQL
Others: Excel, Power Point