WENYI (JULIA) XU

Data enthusiast intrigued to turn data into valuable insights and help businesses grow.
Proficient in data analysis and visualization. Experienced in applying statistical and machine learning models to practical situations.
(646) 881-2612 | wenyi.xu@columbia.edu | Portfolio Website | Github | Linkedin

EDUCATION

Columbia University — New York, NY
M.S. Business Analytics (GPA: 3.83 / 4.00) — Aug 2019 - Dec 2020
- Coursework: Python, R, SQL, Statistics and Simulation, Optimization, Data Visualization, Machine Learning

Shanghai Jiao Tong University (SJTU) — Shanghai, CN
M.E. Electronic Engineering (GPA: 3.50 / 4.00) — Sep 2016 - Mar 2019

Nanjing University (NJU) — Nanjing, CN
B.S. Electronic Engineering (GPA: 4.68 / 5.00, top 5%) — Sep 2012 - Jun 2016

SKILLS

- 2+ years' experiences in Python, R, SQL, Excel; Familiar with Power BI, Tableau, SSMS(Pipelines for ETL), Microsoft Azure
- Proficient in Python Numpy, Pandas, Matplotlib, BeautifulSoup, ML/DL packages including scikit-learn, keras
- Solid knowledge in linear regression, classification, clustering, simulations and hypothesis testing

WORK EXPERIENCE

Panasonic North America — Newark, NJ
Data Scientist Intern / Business Intelligence Group — Aug 2020 - Dec 2020
- Predicted monthly revenue by forecasting transportation time using XGBoost regression model; Kept prediction error under 15%
- Recommended new products based on product specialists' insight and competitors' data scraped by professional tool import.io
- Conducted sentiment analysis on customer reviews from online distribution platforms using Azure Cognitive Service
- Built interactive reports using Power BI and Excel on an ad hoc basis to visualize analysis results

Balyasny Asset Management — New York, NY
Quantitative Analyst Intern / Equity Group — Jan 2020 - May 2020
- Predicted short-term stock price movements based on historical time-series data with a Random Forest (RF) model
- Converted 20 classical trading strategies into quantitative factors; Combined these factors to form new complex factors
- Trained and cross-validated the RF model on a rolling basis; Improved prediction accuracy by 9%
- Extracted a feasible trading strategy from the RF model and proved its superiority by return rate backtesting

Deloitte — Shanghai, CN
Data Scientist Intern / Intelligence Group — Dec 2019 - Jan 2020
- Automated content curation of daily tech newsletter for internal consultants; Saved them over 1 hour's daily reading time
- Performed web scraping and text mining to identify top 10 news topics and corresponding news articles
- Discussed content optimization issues with internal consultants regularly on behalf of intelligence group; Coached new interns

Essence Securities — Shanghai, CN
Research Assistant Intern / Equity Group — May 2018 - Jul 2018
- Wrote investment reports for IC industry with unique insights, garnering views from over 1800 potential clients
- Collected information from over 30 industry experts and company executives through regular in-person conversations

PROJECTS

Small Business Growth in 2010 Winter Olympics (Citadel & Correlation One) — New York, NY
Data Analyst / Team Leader — Jul 2020
- Confirmed structural breaks in 29 regional districts of the British Columbia area via a rigorous time-series statistical test
- Identified sensitive & insensitive industry groups to the Olympic game utilizing spectral clustering algorithm

Squirrel Reporting Web App (CU) — New York, NY
Data Analyst / Team Leader — Nov 2019 - Dec 2019
- Developed a user-friendly Django web app allowing users to report and monitor squirrel sightings in central park area
- Created a PostgreSQL relational database with over 300 records and then connected the web app to this database

Movies' Worldwide Box Office Revenue Prediction (CU) — New York, NY
Data Analyst / Team Leader — Sep 2019 - Dec 2019
- Collected both numerical and categorical features of over 300,000 movies from TMDB web pages using BeautifulSoup package
- Combined internal movie features with external economic indicators to increase prediction accuracy by 5%