WENYI (JULIA) XU

Email: wenyi.xu@columbia.edu | Mobile: (646) 881-2612 | Address: 70 W, 93rd St., New York, NY

EDUCATION

Columbia University — New York, NY
M.S. Business Analytics, GPA: 3.83 / 4.00 (top **10%** of 220 graduate students) — Aug 2019 - Dec 2020
- **Core Courses:** Statistics and Simulation, Optimization, Data Analysis, Machine Learning, Artificial Intelligence

Shanghai Jiao Tong University — Shanghai, CN
M.E. Electronic Engineering, GPA: 3.50 / 4.00 (top **20%** of 72 graduate students) — Sep 2016 - Mar 2019
- **Core Courses:** Data Mining, Database Techniques, Large-scale Data Processing
- **Awards:** 2nd Prize in National Mathematical Modeling Competition, 2nd Prize in Jingmao National Investment Competition

Nanjing University — Nanjing, CN
B.S. Electronic Engineering, GPA: 4.40 / 5.00 (top **5%** of 190 graduate students) — Sep 2012 - Jun 2016
- **Core Courses:** Calculus, Probability, Linear Algebra, Object-oriented Programming
- **Scholarship:** AEGON-Industrial Fund Scholarship, People's Scholarship, Outstanding Student of Nanjing University

SKILLS

- Programming Languages: Python (Numpy, Pandas, Scipy, scikit-learn, Keras, Plotly), R (dplyr, ggplot2), MySQL
- Visualization Tools (Tableau, Power BI), ETL Tools (SSMS, SSIS), Azure (e.g., Databricks, Data Lake, ML Services)
- Data Analytics: SQL, Bloomberg, Wind

PROFESSIONAL EXPERIENCE

Panasonic North America — Newark, NJ
Data Scientist Intern (Business Intelligence Team) — Aug 2020 - Dec 2020
- Built an analytical model to improve the predictability of revenue estimation by segmenting the electronic components based on the product distributors and categories, and by adopting stepwise feature selection to mitigate calculation noise
- Implemented an XGBoost regression model using Python to analyze the impacts of model inputs (inc. manufacturer origin, transportation methods) on the calculation of component delivery time, reducing prediction error from 30% to 15%
- Deployed trained machine / statistical learning models on the Azure cloud platform and automated their dataflows; presented the monthly revenue through a Power BI interactive visualization dashboard

Balyasny Asset Management — New York, NY
Data Analyst Intern (Equity Team) — Jan 2020 - May 2020
- Trained a Random Forest model with the input trading signals such as Moving Average, Bollinger Bands and their combinations to forecast the direction of the daily price movement of Russell 3000 equities
- Conducted stress analysis to understand Covid-19 pandemic impact by identifying the periods of the drawdown and the following recovery period of historical financial crises by comparing the S&P500 Index with VIX Index
- Presented the rationale of the trading strategy and the comprehensive results of the analysis to internal data analysts; gained substantial experience of analyzing large volumes of financial data in accordance with strict client-driven deadlines

Deloitte — Shanghai, CN
Data Scientist Intern (Data Intelligence Team) — Dec 2019 - Jan 2020
- Worked with a team of 4 members to create a digital daily newsletter and automated its content curation process using NLP
- Performed web scraping (BeautifulSoup4) and text mining (NLTK) to identify the top 10 news topics and corresponding news articles; discussed with internal consultants weekly and adjusted curation scripts according to business requirements

Essence Securities — Shanghai, CN
Research Analyst Intern (Equity Team) — May 2018 - Jul 2018
- Conducted detailed fundamental analysis of the Chinese electronic industry based on the supply/demand data and the trend of industry benchmark using public datasets extracted from Bloomberg and Wind
- Built financial valuation models (e.g., DCF) using Excel built-in functions like Pivot Table and Bloomberg Excel Add-In and participated in monthly clients meetings to understand clients' requirements and project deadlines

EXTRACURRICULAR AND INTERESTS

Columbia University — New York, NY
Movies' Worldwide Box Office Revenue Prediction — Oct 2019 - Dec 2018
- Classified over 30,000 movies' profitabilities into 3 levels with 64.6% accuracy based on web scraped IMDB movie data
- Trained Decision Tree, KNN, and Random Forest classification models with both movie data and time-series economic data

Interests: sketch painting, picture editing (Photoshop), video editing