Chen Wang

Cw3211@columbia.edu | (917) 825-7545 | 30-02 39th Ave, Long Island City, NY, 11101

EDUCATION

Columbia University — New York, NY

M.S. in Business Analytics — Aug 2019 - Dec 2020

- GPA:3.92/4.0; Course Assistant for Analytics in Python; Navigator for MSBA Class 2022 & 2023
- Courses: Machine Learning, Simulation, Optimization, Product Management, Analytics on the Cloud

Computer Skills: Python, R, SQL, Tableau, Machine Learning, Django, Scala, Spark, STATA

Certifications: AWS Cloud Practitioner, AWS Solution Architect Associate, AWS Machine Learning Specialty

Central University of Finance and Economics — Beijing, CN

B.S. in Economics, Major in Mathematical Economics and Mathematical Finance — Sep 2015 - Jul 2019

- GPA: 92.27/100; Graduate with Honor (Top 5%); TA for Mathematical Analysis and Advanced Microeconomics
- Courses: Linear Algebra, Probability and Statistics, Econometrics, Stochastic Processes, Differential Equations

PROFESSIONAL EXPERIENCE

Rocket Mortgage (Quicken Loans, LLC) — New York, NY (Remote)

Associate Data Scientist — Mar 2021 – Present

- Lead the winner machine learning project of auto-classification of client relationship emails; Retrieved client relationship email data using **SQL**; Trained **Naïve Bayesian Classifier** (with word-level tf-idf) and got on test set accuracy 0.85, precision 0.82, recall 0.83; Saved 4000+ total working hours each year
- Predicted lead to lock conversion for 2M+ loans in the next 4 months; Extracted CARI lead facts data using **SQL**; Conducted EDA and performed feature engineering using **Pyspark**; Restructured and developed Random Forest classification model with Isotonic Regression calibration in Python on AWS **Sagemaker** hitting test F1 score 0.86; Supported deploying the model in non-prod environment
- Conducted quantitative research supporting Mortgage Servicing Right (MSR) pools with total worth >$40B bidding decisions for Capital Market department; Reformatted the raw data tape; Built reusable solutions to automated the data cleaning process using statistic tools; Devised **Plotly** dashboards with metrics that are critical for decision making

Lenovo | Columbia Capstone Project — New York, NY

People Analytics Intern — Jan 2020 – Aug 2020

- Developed intelligence tool for a 22-person HR global team to conduct sentiment analysis and phrase extraction on any employee reviews, including those of their competitors, contributing to refining recruiting and retention strategy
- Web scraped 100K reviews from Glassdoor, Indeed, Twitter, and Chinese equivalents, of 10 tech companies using **Python** & **R**; used **Clustering** to filter foreign languages; exerted **StandfordCoreNLP** and **Regex** to extract phrases
- Launched visualization and generational analysis of job preferences using **Plotly** based on 300 survey responses designed on **SurveyMonkey** and gathered from **Amazon Mechanical Turk**
- Proposed a complete employee value proposition profile for targeted recruitment of Millennials and Gen Z
- Presented findings and tools to HR managers, who adopted recommendations to emphasize entrepreneurship and teamwork in their Global Future Leaders 2020 recruiting, which receives 1000+ viable applicants worldwide

Office of NYC Comptroller, Bureau of Budget — New York, NY

Data Analyst & Statistician Intern — Jun 2020 – Aug 2020

Columbia Business School — New York, NY

Research Intern — Jun 2020 – Aug 2020

Meituan - Dianping — Beijing, CN

Business Analyst — Nov 2018 - Dec 2018

- Applied **SQL** to query over 60,000 sales records for Kotex and competing brands
- Implemented **Python** to process, analyze, and visualize data to find customers with high value and purchasing frequency for market analysis; built user profiles applying look-alike algorithms to discover features of core users.
- Translated data and KPI's into recommendations for Kotex to improve product selection and marketing, contributing to an online-offline marketing activity which increased the number of stores with shelf–sales by 412% and Gross Merchandise Volume (GMV) by 101%

RESEARCH & PROJECTS

Prediction of NFL Rushing Yards Using Python — Nov 2019 – Dec 2019

- Analyzed a combination of dataset with >500K entries and web-scraped player and team rankings using **BeautifulSoup**
- Visualized the motion of players simulated by uniform acceleration model using **Bokeh** & **imageio**
- Constructed a model with 4 main categories of determinants (team collaboration, player's ability, environment, formation & strategy) and trained model using machine learning algorithms (SVR, RF, KNN and Light GBM)