Tengyu Song

(516)644-3959 | ts3464@columbia.edu | New York
github.com/st3nv | linkedin.com/in/tengyu-song

EDUCATION

Columbia University New York, NY
M.A. in Statistics (GPA:4.17/4.00) Expected, 12/2023

Shanghai University of Finance and Economics Shanghai, China
B.S. in Statistics (GPA:3.44/4.00) 09/2018 – 06/2022

- Honors: Third Prize for People's Scholarship (*Top 10%, 12/2019*)
- Related Coursework: Database, Machine Learning, Time Series Analysis, Data Mining, Text Mining, Statistical Software, R programming, Survival Analysis, Regression Analysis

TECHNICAL SKILLS

SQL, Python, R, SAS, SPSS, MATLAB, C, HTML/CSS, JavaScript, and MS Office

INTERNSHIPS

JD.COM Shanghai, China
Data Application Engineer 09/2021 – 11/2021

- Delved into coupon data on JD.COM from the supply-side perspective. Built a coupon grading algorithm using Gradient Boost Decision Tree to automatically detect and supply good coupons for customers. Boosted 7- day total GMV by 10%.
- Reduced company's average bug response time by 30% by creating a rating system for employee efficiency based on productivity data.
- Performed anomaly detection on the R&D process flow data, combining both pre-made rules and Isolation Forest Algorithm to reach 80% detection accuracy.

Yum China Shanghai, China
Pizza Hut Data Analyst 05/2021 – 08/2021

- Investigated factors that affect customer activation and retention using Random Forest. The insight gained from the model increased reactivation rate of SMS marketing by 3%.
- Processed 2M+ users' purchasing data in the span of three years using Hive and Impala SQL, created pipelines to extract, visualize and analyze customers' buying patterns from data warehouse.
- Applied sentiment analysis to automatically classify user reviews of Pizza Hut on Alibaba based on pretrained natural language processing models.

PROJECTS

How to Effectively Stop False Rumor on Microblog 05/2020 – 02/2021

- Introduced a propagation model to simulate the confrontation propagation process of false rumors and official fact-checking and debunking.
- Crawled over 10,000 tweets related to false rumor and official debunking using python to test how well the propagation model fits the real-world data.
- Developed a machine learning model to predict the impact of false rumor tweets with/without official statements using algorithms such as Random Forest and XGBoost.

Defect Detection of PV Cell Panels 08/2021 – 12/2021

- Utilized Mask-RCNN model in TensorFlow to perform multi-category instance detection on photographed images of PV cell appearance. Achieved 90% recognition accuracy on the testing set.
- Designed a fast-and-precise PV cell image cropping algorithm using OpenCV to enhance the model performance. Significantly reduced the difficulty of identification (Increased AP50 value by 0.1).