Wangrui Hou (Wendy)

Jersey City, NJ | (929)990-5985 | wh916@nyu.edu | linkedin.com/in/wangruihou

Graduating data science master's student with experience in building data ETL & analytics systems and extracting business insights from data. Looking for a data science position that contributes to the data infrastructure and generates valuable insights from customer/product data through advanced data analytics and machine learning.

EDUCATION

M.S. in Data Science, New York University — Sep. 2019 – May 2021

- GPA: 3.875/4.000
- Relevant Courses: Optimization Linear Algebra, Probability & Statistics, Machine Learning, Big Data, Database Systems, Recommender System, Computational Cognitive Modeling

Honors B.S. in Media, Culture, and Communications, New York University — Sep. 2015 – May 2019

- Minors in Mathematics, Psychology
- Magna Cum Laude: 3.832/4.000

SKILLS

Programming | Python (PyTorch, Pandas, Numpy, Scikit-learn, Matplotlib, etc.), SQL, Spark, Tableau, GitHub
Data and Research Skills | Machine Learning, Data Analytics, Data ETL Pipeline, Data Visualization, Verbal & Written Communication, User Research

WORK EXPERIENCE

Data Science Intern, Johnson & Johnson Health Tech — Jun. 2020 – Aug. 2020
Remote from Jersey City, NJ

- Built 3 data ETL pipelines based on Shopify APIs using Python, SQL, and GraphQL on AWS Lambda to help MyStore, Johnson & Johnson's internal e-commerce platform, obtain raw and live data on order history, 47k+ customers, products, and daily inventory status
- Used K-means clustering and CLV modeling to perform customer segmentation based on calculated customer behavior metrics like recency, frequency, monetary value, etc.
- Generated 20+ detailed and interactive Tableau dashboards to visualize data including customer behaviors, revenue history, order volume history, customer segmentations, frequently bought together products, etc.
- Advised J&J Health Tech team leadership of appropriate data-driven business strategies and performed ad hoc data analyses to prepare presentation materials for the strategy team

Client Strategy Intern, Lotame Solutions Inc. — Sep. 2019 – Dec. 2019
New York, NY

- Used Decision Tree and Logistic Regression algorithms for feature engineering on the topic of churn prediction and discovered the important relationship between "Client Touchpoint Score" and client churn
- Advised the Client Strategy team with churn prevention strategies accordingly

Research Assistant, New York University — Jul. 2019
New York, NY

Business Development Intern, Tencent — Jul. 2018 – Aug. 2018
Shenzhen, China

DATA SCIENCE PROJECTS (https://github.com/hwendy12)

End-to-end Spoken Language Understanding System, Deep Learning Project — Sep. 2020 – Present

- Pretrained the system's acoustic branch on LibriSpeech to improve intent predictions with PyTorch
- Modified the model architecture with Python so that it can perform predictions with ASR transcripts as inputs when original audio files are not available due to privacy reasons
- Combined predictions from acoustic and text branches to create better predictions
- Achieved the most state-of-the-art intent prediction accuracy for datasets SNIPS and FSC

Goodreads Book Recommendation System in Spark, Big Data Project — Apr. 2020 – May 2020

- Built an ALS recommendation model with PySpark SQL and MLlib on 2M+ user-item interaction data
- Visualized how interaction data are distributed with UMAP and t-SNE

COVID-19 Infection Rate Prediction, Machine Learning Project — Apr. 2020 – May 2020

- Used Python K-NN to impute and analyzed Johns Hopkins' COVID data and U.S. counties census data
- Trained Gradient Boosting algorithm with Python to predict COVID infection rates for the 529 counties missing from Johns Hopkins' dataset based on their county features data from U.S. Census