Wangrui Hou (Wendy)

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

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, Responsible Data Science, 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
- Relevant Courses: Honors Thesis on Gender Representation Bias in Chinese Advertisements, Algorithmic Culture, Projects in Programming and Data Science, The Rise of Internet Media, Advertising and Society

SKILLS

Programming | Python (PyTorch, Pandas, Numpy, Scikit-learn, Matplotlib, etc.), SQL, GraphQL, Tableau, GitHub, Hadoop, Spark

Skills | Machine Learning Algorithms, Data Analytics, Data ETL Pipeline, Data Visualization, Data Management Platforms, Google Analytics, Verbal & Written Communication, User Research

DATA SCIENCE PROJECTS

See more at: 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
- Modified the model architecture 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 (system combination)
- Achieved the most state-of-the-art intent prediction accuracy for datasets SNIPS, FSC

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

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

Feature Extraction for Successful Restaurants, Machine Learning Project — Nov. 2019 – Dec. 2019

- Trained Decision Tree, Random Forest, and Logistic Regression algorithms to explore key features associated with the restaurants' ratings and numbers of reviews on Yelp

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 performed customer segmentation based on calculated customer behavior metrics like recency, frequency, monetary value, etc.
- Generated 20+ detailed and interactive Tableau dashboards to visualize data insights including customer behaviors, revenue history, order volume history, customer segmentations, frequently bought together products and etc.
- Advised J&J Health Tech team leadership of appropriate 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
- Interpreted first-party and third-party data to access user behaviors/campaign performances and presented these business insights to clients through written reports and presentations
- Communicated with clients on their requests to ensure client satisfaction, engagement, and product usage