YUTING GONG

New York, NY 10023 | 929-310-2608 | yg2641@columbia.edu | http://yutinggong.github.io/

EDUCATION

Columbia University **New York, NY**
M.S. in Applied Analytics, GPA 4.1/4.0 Expected Dec. 2019
- Related Coursework: Python for Data Analysis, Applied Data Science, Managing Data, Storytelling with Data, Research Design

New York University Shanghai **Shanghai, China**
B.S. in Business and Finance, Minor in Mathematics, GPA 3.7/4.0 May 2017
- Honors & Awards: cum laude, Dean's List; Recognition Awards (2014, 2015, 2016), NYU Leadership Fellow
- Related Coursework: Probabilities and Statistics, Linear Algebra, Calculus, Business Analytics, Information System

SKILLS

Data Science Tools: Python (Pandas, NumPy, scikit-learn, NLTK), R (ggplot2, dplyr, Caret), SQL, Tableau, Power BI, AWS, Excel
Algorithms: Linear Regression, Logistics Regression, Decision Trees, Supporter Vector Machine, Clustering, Time Series, NLP

PROFESSIONAL EXPERIENCE

Fractal Analytics **New York, NY**
Data Science Intern May 2019 – Aug. 2019

Predictive Maintenance Project for a global food manufacturing company (Language: Python)
- Effectively featured engineered new variables from sensor data of machines for last 2 years as well as historical maintenance logs and downtimes to predict failures of machines, aiming to reduce opportunity cost of machine downtimes
- Identified the class imbalance problem and used over sampling to improve the prediction precision and recall
- Built a XGBoost classification model to predict machine failures with 90% precision, which helped the company to reduce downtimes

Natural Language Processing Project for a large consumer product company (Language: Python)
- Pre-processed 320,000 rows of unstructured consumer reviews data using NLP tools in Python (e.g., Spacy, NLTK) and transformed the corpora into vector formats such as Bag of Words and TFIDF
- Built an effective n-gram Latent Dirichlet Allocation (LDA) topic model to identify key product pain points from consumer reviews using Gensim; Leveraged model to source root cause of complaints (e.g., product ingredients) to replace manual issue labeling
- Developed actionable insights and making recommendations for the consumer products company to improve certain product features
- Built a web application in Flask for basic NLP tasks as an internal tool for consultants to save 30% of time on coding; the tool prompts users to upload an Excel file of texts, and it automatically processes the texts, runs related models, and returns the result of topics and consumer sentiments in the texts

adidas Group **Shanghai, China**
Rotational Analyst in Business Intelligence, Strategic Planning and Project Management Jun. 2017 – Jul. 2018
- Developed a sales report using Excel VBA/Macros to analyze and visualize sales data in China and identified opportunities for growth in low and high tier cities; Report was distributed to the Head of Performance and Neo brands under adidas
- Evaluated campaign effectiveness by A/B testing, and utilized relevant store metrics to determine effective sales benchmarks
- Used Business Intelligence tools (e.g., MicroStrategy) to build KPI dashboards and prepared bi-monthly adidas China Business Review to consolidate adidas performance trends and financial metrics for managing director's presentation to C-suite
- Partnered with cross-functional teams to develop store segmentation matrix to determine product strategy for retail branches

PROJECTS

Box Office Prediction (Language: R) Mar. 2019 – May 2019
- Analyzed 7,000 observations to predict movie revenues, and recommended ways for movie producers to generate higher revenues
- Imputed missing values, and conducted exploratory data analysis by visualizing correlations among variables using ggplot2
- Conducted feature engineering on numerical variables, as well as text and categorical variables such as "movie overview"
- Trained and tuned machine learning models including SVM, Random Forest and XGBoost, and reached a 2.1 RMSLE

Fast-Food Restaurants Clustering (Language: R) Mar. 2019
- Developed models to cluster fast-food restaurants based on 11 characteristics (e.g., price, food type) using unsupervised learning modeling methods such as hierarchical clustering, K-means clustering with Silhouette method, and model-based clustering

SQL Database Design for Course Managing Data (Language: SQL, NoSQL) Feb. 2019
- Designed a relational database schema in BCNF that enforces functional dependency for Columbia Deli to record employment info
- Created database and populated data using SQL code in MySQL, and generated a view on a set of tables
- Designed and recreated a non-relational database in NoSQL using MongoDB and interacted with the database using PyMongo

LEADERSHIP & ACTIVITIES

Columbia Data Science Society (CDSS) **New York, NY**
Executive Board Member Sep. 2018 – Present
- Moderated Women in Data Science Panel with 5 panelists of female data scientists and started the first mentorship program