MIRANDA ZHOU

EDUCATION

Columbia University New York, NY
Master of Science in Data Science 3.834/4.0 Expected Dec 2021
Coursework: Recommendation & Personalization, Machine Learning, Deep Learning, Databases, Data Visualization

University of California, Berkeley Berkeley, CA
Bachelor of Arts in Data Science, 3.662/4.0, Dean's List May 2020
Coursework: Data Structures, Data Science, Data Mining and Analytics, Casual Inference, Probability and Statistics

SKILLS

Programming	Python, R, Java, SQL
Python Packages	NumPy, Pandas, Scikit-Learn, Keras, Gensim
Databases	MongoDB, Neo4J
Data Visualization	Matplotlib, ggplot2, Seaborn

WORK EXPERIENCE

IBM New York, NY
Senior Data Scientist Intern Jun 2021 - Present

- Utilized supervised and unsupervised machine learning techniques for clustering user personas and determining optimal survey triggering opportunities
- Wrangled 4 million rows of clickstream data stored in company database and visualized 10,000 verbatim survey results
- Presenting recommendations to executives at end of internship, detailing value proposition and strategy for changing to proposed survey method

UC Berkeley Food Pantry Berkeley, CA
Data & Operations Lead Aug 2019 - May 2020

- Led team of 8 to drive operational improvements on key business segments such as client satisfaction, donor outreach, high quantity survey analysis, inventory management, and volunteer shift optimization
- Analyzed over 50,000 student visits with SQL and used tools such as Decision Trees, Logistic Regression, and Times Series Analysis to solve queries developed from stakeholder insights
- Designed relational database for future data collection and data analysis structure of organization

COMET Technologies San Jose, CA
Marketing & Business Analyst Intern May 2019 - Sep 2019

- Improved sales forecasting mechanism by creating a GUI with sales report extraction and error metric analysis
- Developed competitive feature benchmark analysis and price predictors of top 10 competitors by implemented Linear Regression and requested summary statistics while providing clear code to easily add other mechanisms

DATA SCIENCE PROJECTS

Columbia University, MovieLens Movie Recommendation System Jan 2021 - May 2021

- Created end-to-end personalized movie recommendation system through a parallel hybrid model by bucketing users for specific algorithms: Content-Based model, Matrix Factorization, and Wide and Deep model
- Tuned model based on multiple metrics (RMSE, item and user coverage @ k, NDCG, and ranking metric precision recall @ k) for 282,000 users and 54,000 movies and evaluated against a simple popularity baseline model

Columbia University, Detecting Cancer in Pathology Images Sep 2020 - Dec 2020

- Localized cancer cells in 7GB of gigapixel pathology slide images achieving 98% accuracy
- Formulated project as binary classifications of all patches of slide tissue, leveraging insights from combination of two CNNs with different zoom levels as input to determine presence of tumor cells
- Utilized Inception(V3) Architecture for transfer learning and Keras ImageDataGenerator to augment data for robustness