JianHui (Jake) Li

GitHub | LinkedIn | LiJake2001@gmail.com | (917) - 826 - 4237 | Brooklyn, NY

Education

Baruch College, Zicklin School of Business New York, NY

BBA in Computer Information Systems, GPA: 3.8/4.0 (Dean's list 2019 - 2020) Expected May 2023

- **Relevant Coursework**: Data Mining, Big Data Technologies, Analytics Programming, Data Visualization, Database Management, Probability, Statistics, Linear Algebra

Skills and Certifications

- **Languages:** Python, SQL, PySpark, R, Bash
- **Libraries:** NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, Plotly, Tidyverse, ggplot2
- **Framework:** Streamlit
- **Tools:** Git, Jupyter Notebook, AWS, Tableau, Relational Database (PostgreSQL, Microsoft Access), Excel

Certifications: Google Data Analytics Professional Certificate, AWS Academy Graduate - AWS Academy Cloud Foundations, Microsoft Azure Data Fundamentals

Experiences

CUNY Tech Prep / Data Science Fellow Jun 2022 - Present, Virtual

- Learning in-demand technologies such as Python 3, Jupyter Notebooks, Numpy, Pandas, Scikit-learn, and SQL.
- Learning industry best practices for data collection and processing, exploratory data analysis, feature engineering, statistical modeling, data visualization, machine learning techniques, data science process, and big data.

Unadat / Quantitative Analyst Intern Jul 2022 - Aug 2022, Virtual

- Collaborated with a cross-functional team to implement a new feature on the Unadat website and investigated how it can impact and improve customer experience.
- Cleaned and manipulated the survey dataset using R and then used SQL to create score benchmarks.
- Received agile onboarding and scrum methodology training.

Projects

Stock Price Prediction (PySpark) Sep 2022 – Dec 2022, Baruch College

- Built a completed machine learning pipeline using Amazon Web Services (AWS) on a 33GB stock price dataset.
- Built a predictive model using random forest regressor with an r-squared of 0.93 using Pyspark in AWS EMR.
- Efficiently used Boto3 and Python in EC2 to download data directly into S3 and produced descriptive statistics.

Crime Prediction (Python) Sep 2022 – Dec 2022, Baruch College

- Built a predictive model using random forest regressor to predict the number of crimes with an r-squared of 0.93.
- Performed data cleaning, EDA, feature engineering, modeling, and visualizations using pandas, scikit-learn, and plotly.
- Designed and built an interactive web app for users to interact with the model using Streamlit.

Heart Disease Classification (R) Sep 2022 – Dec 2022, Baruch College

- Built multiple classification models to determine the presence of heart disease based on 17 health statuses.
- Used the SMOTE function to oversample the minority class to solve the highly unbalanced dataset issue.
- Performed exploratory data analysis, data restructuring, modeling, and evaluated performance using confusion matrix.

Personal Project

Google Data Analytics Professional Certificate / Bike-Share Case Study Jul 2022

- Identified trends and differences between members and casual riders using historical data and created supporting visualizations to help design marketing strategies for the marketing analytics team.
- Efficiently used R to collect, clean, manipulate, analyze, and visualize data by using tidyverse and ggplot2 in Rstudio.
- Made recommendations based on analysis results to help convert casual riders to members.

Leadership

Technology Leadership Development Program (TLDP) New York, NY

Student Tech Leader Sept 2021 - May 2022

- Selected as one of the 18 students to participate in a yearlong tech leadership development program.
- Attended workshops that improved leadership capacity, professionalism, and career readiness as well as mentorship that broadened my industry knowledge and received career advice.