Jiayi Wang (Skylar)

New York, NY || (646) 891-8987 || w.jiayi@columbia.edu || LinkedIn: www.linkedin.com/in/JiayiWang-Skylar

Data scientist and data analyst with 2-year experience in data analysis, modeling, visualization, and applied statistics. Strong technical knowledge of Python, SQL, Spark, R and Tableau as well as adept in leadership roles

EDUCATION

Columbia University *New York, NY* *Sept 2021 - Expected Dec 2022*
Master of Science in Applied Analytics *GPA: 3.8/4.0*
- Coursework: Machine Learning, SQL and Relational Databases, Python for Data Analysis, Statistics, Data Visualization & Design, Applied Analytics Frameworks, Research Design, Anomaly Detection, Managing Data

The University of Nottingham *Ningbo, China* *Sept 2016 - June 2020*
Bachelor of Science in Economics in the First Class *GPA: 3.8/4.0 (10%)*
- Coursework: Database Design and Implementation, Applied Econometrics, Quantitative Method, Financial Economics

PROFESSIONAL EXPERIENCES

Data Scientist Intern, NSi insurance Group, Miami, FL *Jun 2022 - Present*
- **Improved query efficiency of databases by 15% via implementing sharding**; Designed ETL pipelines for automated database linkage on machine learning models for data wrangling automation [SQL, Hadoop & Spark]
- Created machine learning models for the home insurance comparison search engine to predict customer purchasing probability, delivered customer insights to sales department **with 8% increase in operational efficiency** [Random Forest, XGBoost, H2O]
- Applied exploratory data analysis to generate target consumer profiles for developing the customized marketing strategy [EDA]

Data Analyst, Beauer Tech, Remote *Jan 2022 - May 2022*
- Performed natural language processing on 10,000 consumer reviews to identify market trends and competitive product analysis, developed product optimization strategies [NLP]
- Constructed A/B testing experiments to verify the new strategies effectiveness on user preferences, improved product usage rate
- Created competitive advantages and presented business dashboard to investors to **realize $80,000 investments**

Consulting Intern, PwC Strategy&, Shenzhen, China *Sept 2020 - Dec 2020*
- Developed the Probability of Default model for investors of a P2P lending company to assess counterparties' credit rating
- Implemented logistic regression to build a Personal Lending model mapping the default probability to an ordinal rating system as a predictor for loan interest rate [Python]
- Confirmed model robustness and generalizability through backtesting, benchmarking against alternative models, sensitivity tests, and model assumptions testing [Decision Tree, Random Forest, K-fold Cross-validation]
- Presented the PD model with stakeholders across business verticals and **obtained approval from Model Governance Board**

PROJECT EXPERIENCES

Movie Recommendation Engine Development in Apache Spark *Mar 2022 - May 2022*
- Built an ETL pipeline to analyze movie rating and conducted online analytical processing (OLAP) with Spark SQL
- Implemented ALS matrix factorization on 100,000 ratings to provide personalized movies recommendations and developed user-based approaches to handle system cold-start problem with RMSE 0.79
- Tuned hyperparameters via Spark ML cross-evaluation toolbox and monitored data processing performance via Spark UI on AWS

Bank Customer Churn Analysis and Prediction *Dec 2021 - Feb 2022*
- Developed algorithms to predict customer churn probability via Python and Apache Spark
- Preprocessed data for downstream statistical analysis through imputation of missing values, outlier detection, and categorical feature transformation; Trained Logistic Regression, SVM, Random Forest and XGboost via K-fold Cross-validation [ML]
- Applied regularization with optimal hyperparameter to overcome overfitting and evaluated model performance via F1 score 0.86; Generated Tableau dashboard to visualize prediction results and top factors that influence user retention

Food Search Engine *Sept 2021 - Dec 2021*
- Conducted a food search engine that allows users to explore behind-scene facts of certain food product
- Developed an ETL pipeline to store semi-structured JSON data in MongoDB alongside PyMongo
- Established full-text search functionality with Elasticsearch in combination with MongoDB to query unstructured data sources
- Created a Flask front-end interface to form an API connection with backend data-sources as an interactive, web-based point of contact for search queries

SKILLS

Programming: Python (Scikit-Learn, pandas, numpy, nltk), R, SQL (MySQL), Tableau, PySpark, AWS, MongoDB, Elasticsearch, Google Analytics, Advanced Excel (Pivot Table, V-lookup, VBA)
Methodologies: Decision Tree (Random Forest, XGBoost), NLP (LDA), Time Series, Neural Network, Regularization, PCA, SVM, Logistic Regression; A/B Testing, Causal Inference, Exploratory Data Analysis, Sentiment Analysis, Simulations
Credentials: CFA Level II Candidate, 2nd Place 2021 Smart C-Terminal Science and Technology Competition, 4th Place 2020 Roland Berger Talents Competition (4/1200), Excellence Award 2022 Bain Cup Case Competition