Wes Shi

781-718-8038 | wesshi818@gmail.com | LinkedIn | GitHub

EDUCATION

Columbia University New York, NY — Aug 2022-Dec 2023
MEng Operations Research (**STEM**) - **Analytics** track — GPA 3.88/4.0
Relevant Courses: Stochastic Model, Data Visualization, Optimization, Simulation, Data Analysis, Machine Learning

Brandeis University Waltham, MA — Jan 2019-May 2022
BS Applied Mathematics & Quantitative Economics — GPA 4.0/4.0

Skills

Programming: Python (Sklearn, TensorFlow, Beautiful Soup, NLTK, NumPy, Pandas, Matplotlib, Plotly), MySQL, R, Java
Methods: A/B Testing, Time Series Forecasting, Natural Language Processing, Statistical Modeling, Web Scraping
Tools: AWS S3, Excel (VBA), Hadoop, Word, PowerPoint, GitHub, Git, Redshift, Stream lit, Snowflake, SPSS, Tableau

PROFESSIONAL EXPERIENCE

System2 — Jun 2023-Aug 2023
Data Scientist — New York, NY

- Cooperated with data vendor, optimized and automated ETL process of actively trading stocks' **alternative data (1 TB+)** using **AWS bucket** and redshift database, **reduced extraction time by 60%+**
- Performed quantitative portfolio analytics based on market risk data, adapted **random forest** for selecting influential risk metrics (VaR, Volatility), supported making investment decisions
- Developed 3+ interactive client facing, customizable Stream lit app with automated Python and SQL pipeline for visualizing time series prediction based on **LSTM** and **XGBoost**, achieved MSE of 1.7 on short-term prediction
- Designed and analyzed **A/B test** for clients, evaluated the effectiveness of new credit card offer tag by **T-test**, identified ~**3% growth** on active users
- Assessed investment value of a gym chain by data mining, scraped 8k+ gyms' info, **innovated market growth estimation algorithm** and provided growth strategy insights based on geo-grid data analysis

DG Venture Law Firm — Jan 2023-May 2023
Product Manager — New York, NY

- Confirmed business needs of NDA review product by communicating with legal department, identified 5+ **essential end-user needs** (e.g., cloud-based pipeline), developed detailed functional & development documentation
- Led a team to design and develop an **end-to-end solution** for detecting, highlighting clauses potentially containing 10+ types of deal-breaking information in NDAs, providing modification suggestions
- Implemented text feature extraction with bag of words model and trained classifiers including LinearSVC (f1 ~82.3%), Multinomial Naïve Bayes for clause classification
- Fine-tuned **BERT** model to perform clauses' classification, achieved 95%+ f1 score; utilized corresponding deal-breaking info searching, captured **99%+ improper clauses**, saved **20+** minutes per NDA review

JOANN — May 2022-Aug 2022
Data Scientist — Hudson, OH

- Generated **SQL** commands, and utilized APIs (hdbcli) to automate 1 million+ supply chain data collection from SAP HAP
- Performed **EDA** to the raw dataset using **Python**; identified 14+ types of inaccurate records and automated data cleaning pipeline, which has increased **data integrity by 10%+**
- Constructed **SQLite** table including 280+ ship routes' distance, origin, and destination ports for revamping the original geo-map straight distance, improved feature accuracy for 2 logistics metrics (sea & domestic shipping time) by **40%** and **52%**
- Cooperated with analytics team, applied advanced analytical methods including XGBoost prediction model, achieved model error reduction from 14 to 8.4 days, **improved warehouse efficiency by 30%+**
- Analyzed and quantified the effect of replacing HAP with snowflake, including reducing the query time consuming by **60%+**, presented the insights to stakeholders through Excel dashboards

Capital One — Jun 2021-Aug 2021
Business Analysis Capstone Project — Remote

- Deployed Synthetic Minority Oversampling Technique (SMOTE) based on 790K+ imbalanced data to create synthetic data points, promoted the recall and **addressed class imbalance problem**
- Performed in-sample training, cross-validation, and compared model performances; tuned the hyperparameter using Grid Search by k-fold cross-validation and achieved **75.68% AUC-ROC in identifying the card default cases**

PROJECTS

Netflix Business Analysis Project — Sep 2021-Dec 2021

- Applied numerical algorithms such as SVD and **PCA** to standardize, cluster users based on their browsing history data
- Established Stochastic Gradient Descent model with **L2 regularization** for predicting user rating, achieved loss of **0.28** (0-10)