Juanjuan Song

Austin, TX | 737-202-8880| sjjssl@gmail.com | LinkedIn | GitHub

EDUCATION

MS in Data Science GPA: 3.926 *the University of Texas at Austin* *Expected Dec 2023*

- ***Coursework:*** *Principles of Machine Learning, Design Principles and Causal Inference, Data Structure and Algorithm, Deep Learning, Natural Language Processing*

BS in Mathematics and the Applied Mathematics *University of Science and Technology Beijing*

TECHNICAL SKILLS

- **Programming Language:** Python, SQL, R, MATLAB
- **DS &DA Tools:** Power BI, Data Pipeline, ETL(Extract, Transform, and Load), Hypothesis Testing/ AB Testing, Tableau, Time Series, Spatial Statistics, Causal Inference
- **Machine Learning:** Decision Tree, Random Forest, Naive Bayes, PyTorch, TensorFlow

PROFESSIONAL EXPERIENCE

Data Scientist Intern, Yisen Technology *Jun 2022- Aug 2022*

*Cross-functional Analytics to improve conversion rate and reduce churn rate with **Python***

- Conducted user **segmentation analysis** and **statistical analysis** to identify key features for conversion and presented actionable recommendations to guide marketing strategy, increasing conversion rate by 15%.
- Trained **supervised machine learning** models including **Logistic Regression, Random Forest** on the user data to predict conversion rate in precision 0.99. Built a machine learning **pipeline prototype** to improve efficiency in the business decision-making process by 30%.
- Analyzed the **customer churn** using **survival analysis (lifelines package** in **Python)** to identify at-risk customers and uncovered crucial patterns related to features using **Cox Proportional Hazards model**, provided recommendations to help reduce customer churn rate by 33% and increase the total revenue by 21%.

SELECTED PROJECTS

Movie Review Analytics for Internet Movie Database *Sep 2022*

*A sentiment analysis model for online movie review data using **Natural Language Processing***

- Obtained and processed 50k movie review data points using **Python**(**PyPrind** to visualize the progress and the estimated time until completion).
- Processed text data using **Matplotlib**, **NumPy,** and **NLTK** libraries in Python. Transformed words into feature vectors.
- Assessed word relevancy via **TF-IDF,** trained supervised machine learning models including **Logistic Regression**, **Random Forest** to classify reviews based on their polarity and validated the modeling performance with **ROC**, **AUC** and **confusion matrix** achieved 90%+ accuracy.

Email Marketing Campaign (Python, A/B Test, Time Series) *Nov 2022*

- Performed email multi-arm experiment on 480k users divided into 24 treatment groups to achieve a higher conversion rate.
- Calculated email open rate in treatment groups, obtained the effective emails and identified the best user segment group from the open rate **heatmap** visualization with **Seaborn.**
- Built a data frame including link, activity and conversion rate with Pandas for the 24 treatment groups.
- Conducted one sample proportional **A/B test** on treatment and control groups and concluded the statistical correlation between email campaign strategy and conversion rates.
- Performed **time series** analysis and offered the strategy to increase email sending frequency by 55% to achieve an expected higher open rate.
- Executed a **funnel analysis** of Email Opening→Account Link →Account Fund, identified the biggest decrease along the conversion path and potential improvements, increasing conversion rate by 10%.