XIAOMENG "LAURA" YU

Los Angeles, CA 90007 | (213) 234-8588 | yuxiaome@marshall.usc.edu | https://laurayu0916.github.io

EDUCATION

University of Southern California 08/2021 – 05/2023
Master of Science in Business Analytics GPA: 3.84/4.00
Courses: Data Manipulation, Supervised/Unsupervised Learning Method, NLP, Dynamic Optimization, Database Management

University of Illinois at Urbana-Champaign 08/2016 – 12/2019
Bachelor of Science in Actuarial Science (Minor in Business), summa cum laude GPA: 4.00/4.00
Courses: Statistics and Probability, Life and Annuities, Finance and Investments, Loss Reserve, Risk Management Practices

INTERNSHIP EXPERIENCE

Fashom ***Miami, FL***
Data Analyst Intern 05/2022 – Present

- Trained multi-label and multi-class sequence classification models on reviews using BERT and PyTorch, reached 92% accuracy
- Deployed the trained NLP models to the recommendation engine to decrease the return rate
- Rectified the shirt sizes in customers' profiles based on their purchase history using Matrix Factorization
- Classified customers into different body shapes using K-means clustering
- Built a content-based filtering algorithm based on body shape features using PyTorch to increase sales

Keck Medicine of USC ***Los Angeles, CA***
Business Analyst Intern https://github.com/laurayu0916/Patient-Intake-Workflow 01/2022 – 05/2022
Patient Intake Process Improvement Project Leader

- Scrutinized 250 patient cases from 2020 to 2022 and summarized each case into standard patient intake steps in PostgreSQL
- Calculated average, standard deviation, correlation with conversion for each intake step and visualized the result in Dashboard
- Proposed recommendations to reduce patients' waiting time and improve inquiry-to-patient conversion rate
- Crawled U.S. visa waiting days from travel.state.gov website using Selenium

Tencent ***Shenzhen, China***
Entertainment Market Analyst Intern 09/2020 – 01/2021

- Scraped monthly KPIs of 2,000 mobile games from App Annie and analyzed reasons for KPI fluctuations
- Researched advertising strategies for mobile games based on monthly Share of Voice Rankings
- Studied the difference in number of *Genshin Impact* reviews on social media in U.S. *vs.* Japan
- Presented on *State of Survival* and *Free Fire*, including advertising costs, content marketing, operation objectives, target audience

PROJECTS

Tensor And Matrix Operations (with) PyTorch https://github.com/GitHubOfAndrew/TeAMOTorch

- Collaborate with Andrew to build a Python library, TeAMOTorch, meant to facilitate matrix factorization workflows in PyTorch
- Implement the ranking losses, *e.g.* Weighted Margin-Rank Batch Loss, primarily in use in recommendation systems

NLP Analysis for Dating App Reviews https://github.com/laurayu0916/Dating-App-Reviews-NLP

- Investigated the underlying structure of *Bumble* app reviews and deep dived into *Bumble*'s business profile
- Tokenized reviews with BERT embeddings, clustered similar reviews and derived topics with C-TF-IDF
- Trained RNN and LSTM models to predict important reviews, *i.e.* reviews that would receive lots of Thumb-ups later on

Vehicle Insurance Sales Prediction https://github.com/laurayu0916/Vehicle-Insurance-Sales-Prediction

- Analyzed Health Insurance policyholders' profiles to predict whether they would also be interested in Vehicle Insurance
- Refactored Exploratory Data Analysis code into 5 functions and preprocessed data with PySpark
- Trained and tuned hyperparameters of Random Forest and Gradient Boosted Trees using MLlib with highest recall 0.92

Anomaly Detection for Financial Transaction Data https://github.com/laurayu0916/Fraud-Detection

- Researched transaction data from an e-commerce company to predict fraud transactions
- Extracted time-related features and utilized SMOTE sampling to adjust imbalanced dataset
- Built a scoring system based on predict_prob to decide whether an transaction should be passed, declined or manually inspected

SKILLS

Programming Tools: Python, SQL, R, PySpark, MongoDB, Neo4j
Machine Learning: Random Forest, XGBoost (LightGBM), Neural Network, TF-IDF, RNN, LSTM, Transformers and BERT
Statistical Analysis: Descriptive Statistics, Probability Methodology, Hypothesis Testing(A/B Testing), Time Series Analysis
Certificates: AWS Certified Cloud Practitioner, Tableau Desktop Specialist