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, Property and Casualty, Risk Management Practices

INTERNSHIP EXPERIENCE

Fashom ***Miami, FL***
Data Analyst Intern 05/2022 – 08/2022

- Analyze core data across different stages of the product life cycle and optimize Fashom's operations and analytics
- Perform SQL joins and filters to retrieve relevant data out of database
- Train multi-class sequence classification models on reviews using BERT and PyTorch
- Collaborate with data team to improve the recommendation engine and decrease return rate

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

- Headed Process Improvement Project to reduce patients' waiting time and improve inquiry-to-patient conversion rate
- Scrutinized 250 patient cases from 2020 to 2022 and summarized each case into standard patient intake process
- Calculated average, standard deviation, correlation with conversion of each intake step and visualized the result in Dashboard
- Scraped visa waiting days from travel.state.gov using Selenium Driver in Python

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

- Scraped monthly KPIs of 2,000 mobile games from App Annie and analyzed underlying reasons of performance fluctuations
- Fetched monthly Share of Voice Rankings from 15+ platforms on SensorTower and researched apps' advertising strategies
- Studied the difference in "mentions" on social media of *Genshin Impact* in U.S. (200k+ /month) *vs.* Japan (90k+ /month)
- Presented on *State of Survival* and *Free Fire*, including advertising costs, content marketing, operation objectives, target audience

PROJECTS

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 documents with BERT embeddings, clustered similar documents 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 DataFrame in Databricks
- Trained and tuned hyperparameters of Random Forest and Gradient Boosted Trees using PySpark 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 highly predictive 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

LA Crime Data Analysis and Modeling https://github.com/laurayu0916/LA-Crime-Analysis

- Studied the LA crime data to provide hints on when and how the number of crimes would change in this city via Spark SQL
- Identified top-3 danger areas and visualized the crime events w.r.t category and time to give insights on how to distribute the police

SKILLS

Programming: Python, SQL, PySpark, MongoDB, Neo4j, Tableau, Bootstrap, Advanced Excel
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
Society of Actuaries: Financial Mathematics (FM/2), Probability (P/1)