Luxuan Pan

+1 412-694-4539 | luxuanp@andrew.cmu.edu

EDUCATION

Carnegie Mellon University Pittsburgh | 09/2020 - 12/2022

Master of Information Systems Management - Business Intelligence & Data Analytics (GPA 3.87/4)

Core Courses: *Data Focused Python, Data Structures and Algorithms, Machine Learning, Text Mining, Machine Learning with Large Datasets, Applied Econometrics, Object Oriented Programming in Java, Unstructured Data Analytics*

Tongji University Shanghai, China | 09/2016 - 07/2020

BA, Finance (GPA 91.52), Academic Excellence Awards (top 5%)

TECHNICAL SKILLS

Programming: Python (PyTorch, Pandas, NumPy, Scikit-learn, SpaCy, PySpark), Java, Hive SQL, C/C++, C#, Stata, Excel

Data Analytics: Machine Learning (Logistic Regression, HMM, Neural Network, CNN, RNN, K-Means, KNN, Random Forest, PCA), Causal Inference (A/B Testing, DID, PSM), Sentiment Analysis, Topic Modeling

WORKING EXPERIENCE

Kuaishou Technology | Data Scientist Intern | Department of Advertising Beijing, China | 09/2020 - 07/2021

- **[Modeling and Analytics]** Control Negative Impact of Advertising on DAU in Shopping Festival
  - Implemented an **LSTM model** to predict the negative impact of advertising on DAU as the benchmark to optimize advertising strategies in Python; Customized advertising display frequency based on user activity level to improve user experience; Finally eliminated potential DAU churn by over 70%
  - Identified the type of ads with greatest negative impact by conducting **propensity score matching (PSM)** to match users, contributing to adjusting proportion of different types of advertising and increased retention by 1.2%
- **[A/B Testing Experiments]** Conducted and analyzed 10+ A/B testing experiments on new advertising strategies to evaluate promotion effect (inc. impression, cost, cost per thousand, etc.) and user experience (inc. DAU, app duration, etc.), boosting new advertising revenue by $150k+ daily and improving app duration by 0.2%
- **[Dashboard Design and Monitoring]** Designed 20+ metrics tables and built dashboards to monitor short-video and advertising effect with **Hive SQL**; monitored fluctuations (e.g., cost reduction, surge in churn) using user segmentation; Saved multi-dimensional analysis time by 70%

eBay | Business Analyst Intern | China Analytics Center Shanghai, China | 11/2019 - 04/2020

- **[Sellers Management Analysis]** Collaborated with marketing team to manage and subsidize sellers by evaluating their promotion engagement, number of products listed and sales performance in Hive SQL; increased GMV by 12%
- **[Dashboard Design and Monitoring]** Queried 100k+ sales data and maintained 10+ dashboards to monitor sales performance; Identified top categories, top products and best sellers, and presented insights for improving sales

Henkel | Data Analyst Intern | AG APAC & China Pricing Shanghai, China | 07/2019 - 10/2019

- **[Data Pipelines and UI Platform Design]** Designed pipelines and automotive system of data extraction and cleaning in Python and SQL, reducing 70% manual time; processed 20k+ sales and inventory data and designed a UI platform to approve pricing request using R Shiny

Ipsos | Data Analyst Intern | Social Intelligence & Analytics Shanghai, China | 10/2018 - 12/2018

- Collected customer comments from social media and EC-platforms **using BeautifulSoup** to seek for insights of products; extracted marketing strategies by sentimental analysis; integrated social intelligence reports and recommendations for marketing campaigns

PROJECT EXPERIENCE

County-level Health Factor and Outcome Prediction 12/2021

- Cleaned 2k+ raw data to county level, eliminated outliers with 3-sigma principle and offered statistical analysis; examined features importance for feature selection; applied Logistic Regression and Random Forest on health factors to predict premature death; Optimized hyperparameters through 5-fold cross validation and finally reached 89.8% R-squared

Social Media NLP Analysis on iPhone and Galaxy comments 10/2021

- Crawled 500k+ comments of iPhone and Galaxy on social media; extracted NLP tokenization and performed EDA; compared users' attitudes towards two products by conducting sentimental analysis
- Applied TFIDF and performed LDA to discover 10 topics, visualized topics-comments proportion using T-SNE, helping understand user needs and problems of smartphones