

Yingzhe Jin

yingzhej@andrew.cmu.edu | +1 (718) 801-0788 | [linkedin.com/in/yingzhejin](https://www.linkedin.com/in/yingzhejin)

EDUCATION

Carnegie Mellon University, H. John Heinz III College Pittsburgh, PA
M.S. in PPM - Data Analytics (GPA 3.96/4 with 35% Merit) Sep 2020 - Dec 2022
Courses: Data Structure for Application Programmer, Deep Learning, Advanced NLP, Data Mining, A/B Testing & Analyzing, Machine Learning with Large Dataset

Shanghai University of Finance and Economics Shanghai, China
B.S. in Labor Economics & Social Security Sep 2016 - Jun 2020

TECHNICAL SKILLS

Language: Python (PyTorch, PySpark, Scikit-learn, Numpy, Pandas), R, SQL, PL/SQL, Java, Linux
Data Science: Regression (Linear, Logistic, Lasso, Ridge), Ensemble (Random Forest, GBDT, XgBoost), NLP (Sentimental Analysis, Text Mining, Clustering), PCA, t-SNE, A/B Test, AWS, Databricks
Data Visualization: Seaborn, Matplotlib, PowerBI, Tableau

PROFESSIONAL EXPERIENCE

Kuaishou Technology Beijing, China
Data Scientist Intern - Search Team Jan 2021 - Jun 2021

- Designed **A/B tests** on the single-line mode and conducted in-depth metrics analysis on experiments; the selected UI would increase user downloads by 2M in next 6 months
- Optimized search-related metrics, updated the definition of VLP (videos played longer than 18s) through statistical analysis, and extracted 80k loyal users with a promising retention rate via **Hive SQL**
- Implemented **Random Forest** and **XgBoost models** to extract key factors in 850k new user samples for Aha Incentive program and detected “Aha values” for improving retention rate by over 5% in program groups
- Trained different time series models (**Prophet**, **LSTM**, etc.) on 2-year search-DAU data and established a model pick-up system to fit different periods, boosting the search-page transition rate by 20%

IPSOS Consulting Shanghai, China
Data Analyst Intern - Department of Social Intelligence & Analytics May 2019 - Sep 2019

- Gathered consumers’ comments through 3 social media platforms using **BeautifulSoup** and developed user portrait features, providing companies with user-level segment analysis for further developments
- Collaborated with Data Engineers in developing a comment-sentiment platform for future works
- Analyzed the browsing and shopping behaviors of users and extracted 10 user touchpoints for similar products recommendation and consumer reviews, improving repurchasing rate by 7%
- Consolidated customer feedbacks through word clouds and **Tableau** dashboards during 10+ FMCG projects

RESEARCH & PROJECTS

Application of RNN In Predicting Limit Order Book Dec 2021

- Designed a deep combination of CNN, transform layer, LSTM layer architecture (**DeepLoB**) to predict the middle price movements of the high frequency limit order book (LOB) data of cash equities
- Re-implemented the model via **Self-attention** to address the instability of accuracy (0.76 on avg.)

Scholarly NLP Contribution Graph with RoBERTa Dec 2021

- Replicated the **NLPContributions** results and investigated further by applying existing BERT-based models to improve the performance on sentence classifications with a larger dataset covering more paper topics
- Implemented **RoBERTa** and **SCIBERT** for better classification and increased F1 by 2 through RoBERTa

Speech Predictive Modeling & Data Analytics - Kuaishou Mar 2021

- Collaborated with speech team, cleaned the speech data collected from “voice search”, and divided each search frame into phonemes with 40 parameters through Short-Time Fourier Transform (STFT)
- Applied **Deep RNN** to speech recognition with an eval accuracy of 0.77+ and operated tests on search pages