

# Xuyang Ji

Xj2272@columbia.edu | [GitHub Portfolio](#) | (312) 877-7257 | Chicago, IL 60605

## EDUCATION

### DePaul University

Master of Data Science | GPA: 4.0/4.0

### Columbia University in the City of New York

Master of Science in Strategic Communication

### University of Illinois at Urbana-champaign

Bachelor in Advertising and Public Relations

Chicago, IL

*Expected Grad. Dec 2023*

New York, New York

*Sep 2020 – Dec 2021*

Urbana, IL

*Aug 2016 – May 2020*

## TECHNICAL SKILLS

- **Programming:** Python (Pandas, NumPy, Scikit-Learn, SciPy, Keras, Matplotlib), R (Dplyr, TidyR, Caret, Ggplot2), SAS
- **Data Analysis:** A/B Testing and Experimentation Design, Statistical Inference, Big Data Analytics, Tableau, PostgreSQL, Google Tag Management, Google Analytics 4, Google Ads Measurement, ETL, Advanced Excel
- **Machine Learning:** Supervised/Unsupervised, Regression, Optimization, Pipeline, NLP, Feature Engineering
- **Models:** Linear/Non-Linear, Ensemble methods, Decision Trees, k-NN, Support Vector Machine, Naïve Bayes, Time Series

## ACADEMIC PROJECTS

### Predicting Spending Based on Consumer Analysis [R Studio]

- **Clustering:** Developed Consumer Segmentation portfolio with prototype based on 2,240 consumers across 29 attributes, using KMeans Clustering in conjunction with Factor Analysis for Perceptual Mapping;
- **Predictive Modelling:** Built predictive models to classify unseen customer using kNN and Decision Tree, with parallel hyperparameter optimization and cross validation, achieved an accuracy of 93% with Decision Tree algorithm;
- **Optimization:** Applied Decision Trees with misclassification-based optimization, validating using ROC, Precision and Recall;

### Toxic Comment Classification App with Natural Language Processing Techniques [Python]

- **Ensemble Model:** Developed a comment toxicity detection app with Ensemble Classifier on hard voting, using 3 best performing predictive models including SGD Logistic Regression, Naïve Bayes, and Rocchio Algorithms;
- **Predictive Modelling:** Designed ML pipeline with optimization on predictive models, improved accuracy by 33%;
- **Feature Engineering:** Analysed various NLP methods for feature transformation, including TFIDF and Embedding;
- **Dimensionality Reduction:** Used Latent Semantic Analysis for dimension reduction, hyperparameters are tested with Silhouette analysis and baseline Logistic Regression model;

### Jester Joke Recommender System [Python]

- **Apache Spark:** Developed model-based collaborative filtering model, recommending jokes upon user's request, using Alternating Least Square with Spark, which achieves an RMSE of 4.16;
- Designed and implemented item-based collaborative-filtering with multiple similarity metrics using pipeline.

## PROFESSIONAL EXPERIENCE

### CONTENT SOLUTIONS INTERN

Global Business Marketing, ByteDance Ltd.

Beijing, China

*Oct 2020 – Mar 2021*

- **Conjoint Analysis:** Analysed client surveying data using Multinomial Logistic Regression to estimate product utility scores, identified general optimal asset package using total product utility (TPU);
- **Project Management:** Provided media solutions for clients in cosmetics and Consumer goods; recommended client-specific strategies based on Market Research; increased brand recognition with data-driven narratives by 11%;
- **Data Analytics:** Collaborated with Pricing team to develop processed data template into data normalization process, reducing process cycle time by 65% using PostgreSQL and Advanced Excel; Created monthly selling report using Tableau;
- **A/B Testing:** Fostered product iteration by conducting A/B testing to evaluate product monetization potential, resulting in a 13% increase in CTR; spearheaded new functions training regionally for smooth implementations;
- Contributed to the launch of various global projects with Ads assets for narrative building and campaign monetization, leading to a 13% incremental revenue growth by the end of Q1 2021; developed measure plans to refresh campaign tactics.

### MARKET ANALYST INTERN

Leo Burnett Worldwide

Shanghai, China

*May 2019 – Sep 2019*

- **Data Analysis:** Gather social data from various channels and compiled weekly reports using Tableau; developed testing plans to refresh tactics and optimized conversion rate by 27%; communicated complex data to non-technical stakeholders;
- **Research Design:** Employed research tactics such as focus group, social listening, and A/B testing; developed consumer persona, journeys, and behaviours based on data insights to strategize campaign projects;
- **SAS:** Executed SAS projects on competitor analysis, market trend analysis and forecasting; suggested best practices in support of Client business, and provided both quantitative and qualitative assessment for post-campaign performance;
- Managed media distribution in a timely fashion across various channels by analysing impact of direct communications; collaborated with Creative and Account Executive Teams to drive high-impact results with data insights.