

Dian Yu

📞 858 925 3306 ✉️ dianyu2021@u.northwestern.edu 🔗 [linkedin.com/in/dian-yu-623226139/](https://www.linkedin.com/in/dian-yu-623226139/) 🏠 Evanston, IL

Education

Northwestern University

Evanston, IL

MS, ANALYTICS, GPA 4.0

SEP 2020 - DEC 2021

- Relevant courses: A/B Testing, Big Data, Predictive Analytics, Data Mining, Deep Learning, Databases

University of California, San Diego

San Diego, CA

BA, ECONOMICS, GPA 3.88

AUG 2016 - JUNE 2020

- Relevant courses: Data Analysis, Random Experiment, Statistics, Machine Learning, Algorithm and Data Structures
- Magna Cum Laude; Provost's Honors for 11 consecutive quarters; Phi Beta Kappa honor society membership

Experience

eBay

San Jose, CA / Remote

ANALYTICS INTERN

JUNE 2021 - SEP 2021

- Analyzed eBay's checkout errors which impact more than 2.5 million buyers every month, utilizing SQL and Tableau
- Identified a shipping error display improvement with the potential to add \$1,280,000 to yearly GMV; sized the opportunity by estimating the number of sessions affected and analyzing a similar A/B test's effect on conversion
- Convinced PM partners to prepare A/B tests on the feature change, using data visualizations and impact estimations

McKinsey & Company

New York, NY / Remote

PRACTICUM PROJECT TEAM MEMBER

SEP 2020 - JUNE 2021

- Collaborated with McKinsey data scientists to define corporate financial distress and analyze COVID's impact on it
- Developed ETL pipelines to collect and generate features for 500+ public companies from various APIs using Python
- Applied statistical models to determine feature importance and data analysis to identify patterns and irregularities

Cottage Door Press

Evanston, IL

STUDENT DATA SCIENCE CONSULTANT

SEP 2020 - DEC 2020

- Performed deep-dive analyses on the publisher's Amazon sales and Amazon forecast data with Tableau and Python
- Presented weekly to COO about actionable insights on historical trends and long-term inventory strategies
- Built and fine-tuned a sales forecast model using scikit-learn, lowering Amazon's prediction error by 45%

Alibaba

Hangzhou, China

DATA SCIENCE INTERN

JULY 2019 - SEP 2019

- Improved the performance of an XGBoost revenue forecasting model by 38%, enabling 4 million restaurants and local businesses to make better discount strategies on Alibaba's group-buying platform
- Brainstormed new features and aggregated a total of 91 features from various sources using SQL
- Explored the data with pandas and Matplotlib for insights on feature importance and ideas on data transformations

Projects

Game of Thrones Battle Analysis

github.com/ydDianYu/game-of-thrones-battles-analysis

- Visualized data to find interesting insights from battles in the famous fantasy series utilizing Matplotlib and seaborn
- Identified the best commander, top rivalries, attacker's advantage, and the impact of army sizes on battle outcomes

Tweet as Trump

github.com/MSIA/tweet-as-trump

- Built an interactive Flask application that allows users to post a tweet as if they were Donald Trump and see their predicted number of retweets, utilizing Python, HTML, Docker, YAML, and Makefile
- Processed and cleaned 14,219 of Trump's tweets and trained a bidirectional LSTM model using Keras
- Built the functionality for users to save their input and predictions in an SQLite database or an AWS RDS instance

Technical Skills

- **Languages:** SQL, Python, R, Java, HTML
- **Tools:** Tableau, Hive, Hadoop, Spark, Jupyter Notebooks, PyCharm, Git, AWS, Docker, Flask, APIs, Bash, Excel
- **Data Science:** Data Analysis, Metric Design, A/B Testing, Machine Learning, Deep Learning, Data Pipelines