

ZIHUI ZHOU

601 West 110th Street, New York, NY 10025

(617) 396-6702 | zihui.zhou@columbia.edu | <https://www.linkedin.com/in/zihuizhou/>

EDUCATION

Columbia University

New York, NY

M.S. Operations Research, GPA: 3.83/4.00

Aug 2019 - Dec 2020

Coursework: Business Analytics, Data Analytics, Analytics in Action, Machine Learning, Simulation, Optimization, Recommender System

Boston College

Chestnut Hill, MA

B.A. Mathematics & Studio Art, Mathematics GPA: 3.91/4.00

Aug 2015 - May 2019

Cum Laude, Dean's List First Honor; Pi Mu Epsilon Honor Society

Coursework: Algorithms, Combinatorics, Probability, Mathematical Statistics, Linear Algebra, Computer Science

PROFESSIONAL EXPERIENCE

The Associated Press

New York, NY

Student Data Scientist | Python, Tableau, GCP

Sep 2020 - Dec 2020

- Led a team of five to examine statistical causality (Frobenius norm analysis, lag correlation analysis) between B2B and B2C platforms based on different categorical variables, and increased news coverage efficiency by 36%
- Deployed K-means clustering on 500+k articles based on topic modeling (LDA, Non-negative Matrix Factorization with tf-idf)
- Delivered and maintained dashboard of news trends through Tableau, and reported potential profitable topics for news coverage of B2B platforms based on 8M time series data, which increase B2B clients by 12% and total profits by 15%

Intellipro Group Inc.

Santa Clara, CA

Data Analyst Intern | SQL, Python, Tableau, JavaScript

May 2020 - Aug 2020

- Identified talent flow in social media industry and generated data-driven recommendation for talent management by scraping, analyzing, and visualizing over 10k data points through JavaScript, Python, and Tableau
- Modeled employee churn with 56% improvement through implementing KNN, Random Forest, SVM, Logistic Regression
- Conducted sentiment analysis and LDA topic mining on 10k data to optimize employee retention problems

BrainCo Inc.

Somerville, MA

Research Assistant Intern | MATLAB, Python, Excel

Mar 2018 - Sep 2018

- Collected, cleaned, and visualized 357 Electroencephalogram (EEG) signals through Python, MATLAB, and Excel
- Compared, classified, and identified EEG datasets of normal children and children with Autism Spectrum Disorder using Principal Component Analysis, SVM, and Short-Time Fourier Transform with 32% improvement on accuracy in a cross-functional setting

ACADEMIC PROJECTS

Instagram Employee Churn Prediction

Jul 2020 - Aug 2020

- Collected, preprocessed, and cleaned 200+k Instagram employee data and relative features in professional experience, education level, and skills through LinkedIn API and web scraping
- Executed data cleaning and extracted topics through text mining (NLP) using NLTK and Pandas
- Built binary classification models (Logistic Regression and Random Forest) and improved accuracy by 23% through cross-validation

Sentiment Analysis of Zoox Buyout 2020 Based on Reddit

Jun 2020 - Jul 2020

- Scripted Reddit API and managed MongoDB to analyze over 10,000 captured posts related to Amazon's acquisition of Zoox
- Applied Google Cloud Natural Language API to analyze Zoox's development and the public's attitude towards the purchase through attitude score and magnitude; gave clients recommendations on stock purchase which prevented over \$1M loss

Business Recommender System based on Yelp dataset

Sep 2019 - Dec 2019

- Led a team of three to analyze 6.7 million reviews over 1.6 million users across 192+k businesses
- Reduced MSE by 64% using an ensemble model (Collaborative Filtering, Factorization Machine, Text Mining) formed linear regression of log-return overtime on 10 selected consumer staples industries' stocks over 2 years

SKILLS

- Software: Python (NumPy, SciPy, Scikit-learn, pandas, Matplotlib, Plotly, NLTK, Spark, TensorFlow, BeautifulSoup), SQL, Tableau, Excel, MATLAB, BASH, Django, Git, Spark, JavaScript, HTML, CSS, Swift
- Techniques: Text Mining, Web Scraping, Exploratory Data Analysis, A/B Testing, Network Analysis, Data Visualization