

Zheng Lu (Alice)

New York, NY * (310) 745-6425 * zl3388@columbia.edu * [LinkedIn](#) * [GitHub](#) * GHC2023

EDUCATION

Columbia University

Sept. 2023 – Dec. 2024

Master of Science, Applied Analytics

- Relevant coursework: Data Visualization & Design, Machine Learning, Applied Text and NLP, Anomaly Detection, Data Analytics Using SQL and Relational Databases.

University of California, Los Angeles (UCLA)

Sept. 2019 – Jun. 2023

Bachelor of Science, Statistics and Data Science

- Relevant coursework: Computation with R, Python with Application, Probability, Monte Carlo Methods, Statistical Model in Finance, Time-Series Analysis, Data Analysis and Regression, Large Scale Data Mining, Data Science, Optimization, and Machine Learning.

SKILL

Technical: R, SQL, Python (NumPy, Pandas, Matplotlib, Scikit-learn), Tableau, Power BI, ETL, Pivot Table.

Methods: Prob. & Stats, A/B testing, Experimental Design, Visualization, Machine Learning (Regression, Clustering, Classification, Segmentation).

Leadership: Chinese Scholars and Students Association (CSSA) President, Student Body Senate.

WORK EXPERIENCE

Allied Millennial Partners, LLC

Jun. 2022 – Sept. 2022

Financial Data Analyst Intern

New York

- Established end-to-end ETL pipeline on Tesla stocks and market index data and summarized statistical inference. Constructed an ARIMA model with an AIC -298.78 by time series theory.
- **Optimized final model accuracy in 17%**, leveraged test sets to make predictions with a DW statistic around 2, and reported the final data analysis.
- Reconciled data across 5+ relational databases of electric vehicles (EVs) and enhanced operation efficiency.
- Led a team in Lucid company data analysis, joined company data with corresponding industry, identify investment opportunities and traps by Power BI.

Huawei Technologies Co., Ltd

Apr. 2021 – Aug. 2021

Data Analyst Intern

Beijing

- Designed open-source events database by MySQL, revised retrieval efficiency by 30% by sorting out media conferences and integrating project statistics.
- Collaborated with Huawei Developer Cloud 2021 operation team, reviewed customer profiles, and proposed marketing plans to **overhaul Search Engine Optimization (SEO) by 25%**.
- Developed media data platforms operation, and produced monthly media operations data reports in Tableau.
- Independently responsible for "Customer Big Data Insight Report" of new product M5 (EV model with Harmony OS), tracking metrics and showing a summary of data on key points of products.

The Last Mile Entertainment (TLME)

Jan. 2020 – Jun. 2020

Cofounder/Business Analyst

San Francisco

- Analyzed ticket rate data by MySQL and customer profiling, designed target publicity campaigns, and **increase attendance by 35%**.
- Formulated A/B testing, utilized Python Pandas to analyze user distribution and behavior increased 5% conversion rate.
- Extracted and merged key metrics by pivot table, improved tuning parameters by 10% in data loading efficiency.
- Performed industry analysis and market conditions by SWOT, produced Chinese rapper's tour in China and U.S.

PROJECT

Classification Users' Intention to Purchase New Energy Vehicles

Feb. 2023 – Mar. 2023

- Assessed TF-IDF method to evaluate text importance and filter text. Operating LSI for **text feature extraction** and data dimensionality reduction.
- Data cleaning by filling in missing values, encoding purchase intention levels, boost samples by SMOTE.
- Constructed **Pipeline Grid Search** to settle best combination of optimal feature extraction method (Lemmatization), dimensionality reduction (LSI, k=80), and classifier model (SVM). Final model obtained **86% accuracy**.

Regression Analysis for Vehicle Characteristics

Dec. 2022 – Jan. 2023

- Conducted thorough statistical analyses on vehicle characteristics, such as make, model, engine size, and features, to identify relationships and relevance to vehicle prices. Worked with large datasets to ensure provided an accurate representation of findings.
- Contemplated Random Forest algorithms to determine most effective combination to predict vehicle prices. Deployed various parameters to fine-tune model with maximum depth 4 and obtained an **Out-of-Bag Error of 0.023**.
- Conducted a theoretical and real-world comparison of Polynomial Regression, Linear Regression models with and without Lasso, Ridge penalty on vehicle data sets and advanced prediction to 81% R2.

Improve Electrical Vehicle Battery Life Using Machine Learning Model

Sept. 2022 – Dec. 2022

- Imputed missing value by KNN and one-hot encoding on drivers' personality. Extraction and optimization of driving behavior feature parameters based on full data by PCA.
- Classified driving behavior by dividing it into 5 levels from mild to intense by K-means in R.
- Tested advantageous: random forests, XGBoost, Naïve Bayes, and SVM in training set. Used random forests as final model and **achieved 83% accuracy** in predicting battery life with testing data.

Collaborative filtering model to build a movie recommendation system

Sept. 2022 – Sept. 2022

- Suggest user-item interactions through two algorithms (Neighborhood-based and Model-based).
- Build a model based on a User-based algorithm, design a k-NN collaborative filter, use 10-fold cross-validation to evaluate its performance. By comparing RMSE and MAE values, set best k=12 with 75% accuracy.
- Establish latent factor model (LFM) and collaborative filter of NMF, MF, and Naive models. Find best MF Models with Latent Factor 8 **with 0.78 precision score**.

EXTRACURRICULAR INVOLVEMENT

Auto racing; tennis; badminton; hiking.