

Shenghua (Isaac) Zhu

shenghua.zhu.2023@anderson.ucla.edu | [linkedin.com/in/isaac-zhu](https://www.linkedin.com/in/isaac-zhu) | (424) 324-0541 | Los Angeles, CA, 90024

EDUCATION

UCLA ANDERSON SCHOOL OF MANAGEMENT

Los Angeles, CA

Master of Science in Business Analytics (MSBA – F1 STEM OPT)

Expected December 2023

Coursework: Advanced Machine Learning, Customer Analytics, Operations Analytics, Database Management, Prescriptive Modelling, Competitor Analytics, Data Storytelling, Product Management, A/B Testing

UNIVERSITY OF CALIFORNIA, LOS ANGELES (UCLA)

Los Angeles, CA

Bachelor of Science, double major in Applied Mathematics (Specialization in Computing) and Statistics

Cumulative GPA: 3.86/4.0 (Cum Laude, Dean's Honor List)

Sep 2018 - Jun 2022

Coursework: Data Structures (C++), Algorithms, Probability Theory, Mathematical Modelling, Optimization

TECHNICAL SKILLS

Languages: SQL, Python (PyTorch, Scikit-learn, Pandas, NumPy, Seaborn, PySpark), R, Java, C++

Software: Tableau, Power BI, Apache (Hive, Spark, Hadoop), Git, MYSQL, Excel, PowerPoint

PROFESSIONAL EXPERIENCE

TENCENT

Shenzhen, China

Data Scientist Intern

Jun 2021 - Sep 2021

- Constructed automated data pipeline in PySpark, recording search-related data and reducing querying time by 70%
- Cleaned and analyzed 10 million+ entries of keyword search and user behavior data in Spark, facilitating optimized searching and recommendation algorithms while increasing search accuracy
- Developed Gross Merchandise Value (GMV) attribution models by applying logic trees, Jensen-Shannon divergence and Pareto analysis with over 80% accuracy to detect GMV anomaly and fluctuation

GENETSIS E-COMMERCE

Shanghai, China

Data Scientist Intern

Jun 2020 - Sep 2020

- Created real-time sales and revenue dashboards in PowerBI based on market data for cosmetic industry clients, which was developed into one of Genetsis' essential data products, helping to grow department revenue by 8%
- Initiated and performed pricing analysis on 5K+ cosmetic SKUs with ridge regression models to predict profit maximizing prices, achieving a 0.38 RMSLE, and allowing retailers to make data-driven pricing decisions
- Refined database design in Hive, enabling faster read speeds and reducing dashboard load times by 35%

ANALYTICS PROJECTS

Federated Learning for Data Privacy in Electronic Medical Records (UCLA)

- Created a communication framework for client-server interaction and implemented federated learning on it
- Validated federated learning model on the MIMIC-III Clinical Database to predict patients' mortality with 82% accuracy and 0.86 AUC and reflected on possible factors that contribute to lower mortality rates
- Compared model performance with centralized models and checked model generalization on different data samples

Prediction of Glass Properties with Data Analysis and Machine Learning (UCLA PARIS Lab)

- Cleaned raw datasets of more than 100,000 entries with standard data cleaning pipelines and various outlier detection algorithms, leading to a 75% reduction in errors and improving model accuracy by around 28%
- Utilized classification, clustering, and regression analysis to predict different properties of glass samples like Young's modulus and Poisson's ratio, which helps find the desirable chemical forms of glass materials

Analysis of Percentage Change in YouTube Video Views (Kaggle)

- Performed feature selection on 300+ predictor variables with boosting and lasso using gbm and caret libraries in R
- Coordinated a team of 4 in model selection and building and predicted video growth rates by constructing a random forest regressor in Scikit-learn with an accuracy score of 0.87, which took the lead on Kaggle's public leaderboard