YIHANG HUANG https://www.linkedin.com/in/yihang-huang-367769286/
San Diego, CA (619) 953-8793 yihang.huang@rady.ucsd.edu

EDUCATION

Master of Science in Business Analytics, Rady School of Management 12/2025
University of California, San Diego, CA

- **GPA:** 4.0/4.0
- **Relevant Courses:** SQL & ETL, Statistic NLP, Web Mining & Recommender Sys, Scalable Analytics

Bachelor of Science in Mathematics 06/2024
Shanghai University of Finance and Economics (SUFE), Shanghai, China

- **Awards:** Third Prize of People's Scholarship (2020 & 2021), Outstanding Student of College (2020)
- **Relevant Courses:** Probability and Statistics, Big Data Processing, Stochastic Process Theory, Econometrics
- **Leadership:** Minister of the Student Liaison Department in School of Mathematics

SKILLS

- **Programming:** Python (NumPy, Pandas, Scikit-learn, PyTorch, NLTK), R, MATLAB
- **Big Data & Cloud:** ETL processes, Spark, SQL (MySQL, Snowflake), AWS, Google Analytics
- **Business Intelligence & Visualization:** Power BI, Tableau, Google Sheets, Excel (Power Pivot, Power Query)
- **Statistical Methods**: A/B Testing, Hypothesis Testing (t-test, Chi-squared), Regression Analysis, Bayesian Inference

EXPERIENCES

Data Science Intern, Henkel Investment Ltd., Shanghai, China 07/2023 - 05/2024

- Developed **SQL-based ETL pipelines** to extract, clean, and process 200K+ daily records from SAP, ensuring consistent dealer data across multiple datasets and improving data integrity by 40%
- Automated data consistency checks, implementing **logistic regression model** to flag 2% of high-risk records, and integrated real-time updates in **Power BI**, enabling seamless monitoring data quality
- Built an **LSTM-based forecasting model** to anticipate dealer growth, reducing forecasting errors to 1% and enhancing inventory planning and market expansion strategies
- Built interactive **Power BI dashboards** with advanced **DAX functions** to visualize forecasted trends, calculate deviations, and support real-time monitoring of key metrics, enabling informed decision-making across teams
- Collaborated with IT department on a data migration project, implementing **A/B testing** to validate performance metrics through **t-tests**, leading to a 35% improvement in evaluation efficiency

Project Management Intern, Ernst & Young Corporate Advisory Ltd., Shanghai, China 12/2022 - 03/2023

- Analyzed operational data using **SQL** to monitor KPIs for 30+ projects weekly, identifying underperforming initiatives and reducing costs by 60%
- Implement **ARIMA model** to forecast actual expense data for first half of 2023, ensured model accuracy exceeded 99%, contributing to more effective expense control and financial planning
- Performed **sensitivity analysis** via **Monte Carlo simulations**, modeling the impact of cost drivers (labor and operational volumes) for scenario-based expense optimization

PROJECTS

Predicting User Ratings Using Hybrid Textual Feature Models, Web Mining & Recommender Systems, UCSD 12/2024

- Teamed with students from CS major on a project to predict user ratings by applying a **Latent Factor Mode**l with Singular Value Decomposition (SVD), optimizing hyperparameters to achieve an MSE of 0.9571
- Addressed the cold-start problem by developing a hybrid model that combined SVD latent factors with textual features **TF-IDF**, integrating them through **Ridge Regression**, which reduced the MSE to 0.5290
- Improved model generalization for sparse datasets by exploring **Factorization Machines** (FM) and incorporating **Word2Vec** embeddings and one-hot encoding for user and item features

Chinese-to-English Translation Based on LLM Hypotheses Integration, Statistic NLP, UCSD 11/2024

- Developed a hypothesis integration strategy based on **mBART**, leveraging the top-N outputs of **beam search** and integrating them through **fine-tuning LLM models** such as **LLaMA**, achieving a 0.648 BLEU score on the test set
- Designed cross-model pipelines for initial translation and hypothesis integration, analyzing performance trade-offs in terms of BLEU scores, training time, and computational efficiency