Yutian Yang

501-368-9640 | charlieyang990314@gmail.com | charlieyang1557.github.io/aboutme/ | linkedin.com/in/yutianyang

Education

Trine University • *Master of Science* • Business Analytics 01/2024 – 05/2025

University of California, Davis • College of Letters and Science 09/2021 – 06/2023
Master of Science • Statistics

University of California, Davis • College of Letters and Science 09/2017 – 06/2021
Bachelor of Science • Statistics | Economics

- **Relevant Coursework:** Advance Statistical Computing, Algorithm Design & Analysis, Econometrics, Optimization of Big Data Analytics, Statistical Machine Learning I, Statistical Methods of Machine Learning, Time Series Analysis, Probability Theory.

Skills & Tools

- Programming Skills: Proficient in SQL, Python (Pandas, Scikit-learn, TensorFlow), R
- Data Science Tools: Github, Google Cloud Platform (GCP), Google Analytics, Selenium, Tableau, LaTeX, AWS, Power BI

Work Experience

Data Science Intern – Pinecone 06/2024 – 08/2024
New York, NY

- Designed and built the Book of Business and Account 360 dashboards using **SQL** and **Sigma**, improving sales operations by 15%. Implemented Row-Level Security (RLS) for tailored views, and documented processes in Notion, reducing onboarding time by 30% and ensuring consistent use across teams.
- Developed the "dim_assistants" schema and implemented it in the pipeline using **BigQuery** and **DBT**. Created the Pinecone Assistant dashboard using **SQL** and **Sigma**, enabling comprehensive tracking of metrics. Facilitated cross-team collaboration, leading to a 25% increase in product insights.
- Conducted churn analysis using **Python**, identifying 5 key metrics and setting up alerts, reducing churn by 10%. Overcame data limitations and improved data collection, projected to boost accuracy by 20%.

Data Analyst Intern – Allschool 06/2022 – 08/2022
San Mateo, CA

- Enhanced impression targeting strategies and boosted customer engagement through **A/B testing** and segmentation analysis of user traffic and revenue across regional and platform data utilizing **Google Analytics**.
- Evaluated user behavior across multiple advertising channels, leading to a 15% reduction in project budget and an increase in daily active users employing **SQL** and **BI tools**.
- Designed a real-time web scraper with **Python** and **Selenium**, accelerating the class selection process by 50%.
- Developed a key metrics dashboard for active users, daily traffic, and revenue, improving business visibility and supporting data-driven decision making using **Google Looker Studio**.

Research Assistant – UC Davis Department of Economics 07/2020 – 09/2020
Davis, CA

- Analyzed behavioral trends in procrastination and present-biased behavior using a **generalized linear model (GLM)** and **Logistic Regression** in a study with Professor Anujit Chakraborty.
- Enhanced the reliability of study results by employing **Bootstrapping** resampling techniques to expand the sample size to approximately 20,000 data points.
- Addressed multicollinearity among predictors and improved prediction accuracy of procrastination behavior variables with regularization using **Lasso Regression**.
- Facilitated industry application by uncovering procrastination patterns, offering insights for tech companies to develop user-centric products and services, potentially enhancing user satisfaction, retention, and success.

Data Science Intern – Launchpad Project Management 06/2019 – 09/2019
Davis, CA

- Led and leveraged market analysis and managed databases to construct predictive models that significantly enhanced real estate investment strategy, utilizing **R** and **SQL**.
- Analyzed survey data, extracting critical insights, which led to an estimated 10% increase in investment returns by applying machine learning algorithms and statistical methods like **generalized linear regression** and **logistic regression**.
- Enhanced strategic decision-making with advanced data visualization in **Tableau**, improving team understanding and engagement.

Projects

Classification of Mushrooms: Edible or Poisonous (Python) UC Davis - STA 221 Advanced Statistical Computing

- Collaborated on a team project developing machine learning and deep learning models, including **Random Forest**, **Kernel SVM**, and **Convolutional Neural Networks (CNN)** to classify mushroom images into edible or poisonous.
- Applied advanced techniques such as grid search for hyperparameter tuning and transfer learning using pre-trained ResNet50 model to improve the efficiency and accuracy of the classification models.

- Critically evaluated model performance, suggesting potential enhancements by exploring alternative pre-trained models or architectural adjustments.