

YUTIAN YANG

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

EDUCATION

- 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: SQL, Python, R
- Machine Learning: Scikit-learn, TensorFlow, Random Forests, SVM, CNN, Time Series
- Data Engineering: BigQuery, DBT, AWS, GCP, ETL, Schema Design
- BI & Analytics: Tableau, Power BI, Looker, Sigma, Google Analytics

WORK EXPERIENCE

- Onshape by PTC** (Cloud-Based CAD Platform) – Data Science Intern 06/2025 – 08/2025
Boston, MA
- Built and evaluated multiple unsupervised anomaly detection models (Prophet + Isolation Forest, Merlion, LSTM-AE) for API telemetry monitoring; selected and deployed Prophet-based pipeline for **best performance, interpretability, and ease of deployment**.
 - Applied AWS Bedrock's Titan Embeddings and Claude 3.5 Sonnet for clustering, naming, and sentiment analysis of NPS feedback, **extracting key themes to support product insights** from unstructured text.
 - Conducted keyword analysis on AI Advisor open-field queries to **assess reference URL coverage, uncover content gaps, and improve user-facing query resolution**.
 - Developed Looker dashboards to track anomaly alerts and AI-driven insights, **enhancing visibility across teams** and automating Slack-based reporting to **reduce manual monitoring efforts**.
- Pinecone** (AI Vector Database Unicorn) – Data Science Intern 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 and Random Forests, 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%**.
- Allschool** (EdTech Platform) – Data Analyst Intern 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.
- UC Davis Department of Economics** (Data Analysis & Modeling) – Research Assistant 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**.

PROJECTS

- Classification of Mushrooms: Edible or Poisonous (Python)** UC Davis - STA 221 Advanced Statistical Computing
- Developed Random Forest, Kernel SVM, and CNN models with grid search hyperparameter tuning and **transfer learning using pre-trained ResNet50** to classify mushroom images as edible or poisonous.
- Resume RAG Chatbot** [Live Demo] Python, React, FastAPI, FAISS, Anthropic Claude API
- Built a RAG chatbot with **hybrid retrieval** (BM25 + FAISS), Claude LLM streaming, and query rewriting. Alpha sweep via LLM-as-judge found **pure semantic retrieval achieved 5.0/5.0 relevance**.
 - Deployed full-stack app (React + FastAPI) with source attribution, dark/light mode, rate limiting, and **Docker containerized one-command startup**.