

YUTIAN YANG

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

SUMMARY

Actively seeking opportunities in **DA, DS, and MLE**. I bring proficiency in **Python, R, SQL**, advanced **statistics**, and **machine learning**. I offer a proven background in **data analysis** and **visualization**, with experiences including **CNN-based mushroom classification**, housing price prediction via **GLM**, and the creation of key metrics dashboards using **SQL** and **BI** tools.

EDUCATION

- | | |
|--|-------------------|
| Trine University • <i>Master of Science</i> • Business Analytics | 01/2024 – 12/2025 |
| University of California, Davis • College of Letters and Science
<i>Master of Science</i> • Statistics | 09/2021 – 06/2023 |
| University of California, Davis • College of Letters and Science
<i>Bachelor of Science</i> • Statistics Economics | 09/2017 – 06/2021 |
- **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:** (5+ yoe) 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 Analyst Intern – Education Services Organization
Allschool Inc., Remote | 06/2022 – 08/2022 |
|--|-------------------|
- Utilized **Google Analytics** for a data-driven analysis of regional and platform-specific user traffic and revenue, leading to optimized impressions targeting and increased customer engagement.
 - Leveraged **Python** and **Selenium** package to design a real-time web scraper, providing centralized class schedules and **accelerating the selection process by 50%**.
 - Employed **SQL** and **BI tools** to evaluate user behavior across multiple advertising channels, informing strategic decision-making and resulting in a **15% reduction** in project budget and an increase in daily active users.
 - Contributed to the development of a key metrics dashboard highlighting active users, daily traffic, and revenue, thereby improving business visibility and supporting data-driven decision making.
- | | |
|--|-------------------|
| Research Assistant – Academic Research
UC Davis Department of Economics, Davis, CA | 07/2020 – 09/2020 |
|--|-------------------|
- Utilized **generalized linear model (GLM)** and **Logistic Regression** to analyze behavioral trends in procrastination and present-biased behavior in a study in collaboration 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**.
 - Improved prediction accuracy of procrastination behavior within the test dataset by addressing multicollinearity among predictor variables 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 Analyst Intern – Technology Consulting
Launchpad Project Management, Davis, CA | 06/2019 – 09/2019 |
|---|-------------------|
- Utilized **R** and **SQL** for comprehensive market analysis and database management across multiple portfolios, enabling the development of a predictive model for real estate investment strategies.
 - Applied machine learning algorithms and statistical methods like generalized linear regression and logistic regression to analyze survey data, extracting critical insights, which led to an estimated **10% increase** in investment returns.
 - Employed advanced data visualization tools like **Tableau** to present data analysis results, facilitating strategic decision-making and enhancing team understanding.

PROJECTS

- | | |
|--|--|
| Impacts of Vaccination Policy on COVID-19 Cases (R) | UC Davis - STA 237A Time Series Analysis |
|--|--|
- Teamed up to investigate the impact of government vaccination policies on COVID-19 cases across selected countries using **ARIMA** and **ARIMAX** time series models.
 - Managed data preparation processes involving merging tables, data filtering, and the creation of a "weekly contrast" variable using "Our World in Data".
 - Leveraged **data visualization in R** to effectively communicate findings and performed inferential analysis on model predictions, contributing to a broader understanding of policy effectiveness.
- | | |
|--|---|
| 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 through the exploration of alternative pre-trained models or architectural adjustments.