Yama (Ya-Wen) Chang

Tel: 646-927-9876 | Email: yamachang.ds@gmail.com | LinkedIn | Portfolio | New York, NY

EXPERIENCE SUMMARY

Data Scientist with over 3 years of experience driving data-driven insights, specializing in experimental design, machine learning, and predictive modeling. Proficient in R, Python, SQL, and data visualization tools with a proven track record of delivering actionable insights.

PROFESSIONAL EXPERIENCE

Data Scientist, *Lab for Scalable Mental Health* | *Stony Brook, NY* ***August 2022 – Present***

- Developed a machine learning model for depression prediction using time-series data, resulting in a 20% improvement in diagnostic accuracy and a 15% reduction in misdiagnosis rates.
- Achieved a 90% increase in solution efficiency through the implementation of analytical tools and automation.
- Enhanced engagement rates by 35% and improved communication skills by 40% by identifying key metrics and conducting rigorous exploratory and statistical analysis to revamp a staff communication training program.
- Spearheaded best practices for open data management using GitHub and Open Science Framework, including data quality checks, cleaning, dictionary creation, and statistical analysis to promote transparency and data integrity.

Senior Data Analyst, *University of Pittsburgh Medical Center* | *Pittsburgh, PA* ***June 2020 – July 2022***

- Automated data aggregation from multiple sources, leading to a 20-hour monthly reduction.
- Created a markdown-based reporting system for visualizing recruitment indices, saving 20 hours per month.
- Designed and implemented research studies of survey and behavioral data, yielding critical insights into target demographics and leading to published papers in high-ranking journals.

TECHNICAL SKILLS

- **Programming:** R (tidyverse, ggplot2, tidymodels, markdown), Python (numpy, pandas, scikit-learn, matplotlib)
- **Tools:** SQL, Qualtrics, Git, GitHub, Command Line
- **Stats & Machine Learning:** Covariance and correlation modeling, Hypothesis Testing, A/B Testing, Multilevel modeling, Factor Analysis, Principal Component Analysis, Cluster Analysis/Latent Profile Analysis, Classification (KNN, Random Forest, XGBoost), Regression (Lasso, Ridge, Elastic Net), Clustering (K-means, GMM) Modeling

SELECTED PROJECTS

Real Estate Prices Prediction ***2023***

Machine Learning Regression | *Python*

- Developed a predictive model using machine learning algorithms, including Regression Models, Random Forest, and Gradient Boosting Decision Trees, to estimate real estate prices.
- Optimized model performance through cross-validation and hyperparameter tuning, achieving a Mean Absolute Error under the pre-set benchmark ($70k), reflecting outstanding prediction accuracy.

Geospatial Analysis of Discrimination in the U.S. ***2023***

Factor Analysis & Regression Analysis | *R*

- Extracted, cleaned, and analyzed large open datasets (600000+ raw data) to conduct a comprehensive analysis of attitudes towards LGBTQ+.
- Created and visualized a county-level discrimination index using principal component analysis.

Suicidal Profile Classification ***2022***

Cluster Analysis | *R*

- Utilized cluster analysis to identify distinct profiles over time in elderly individuals with major depressive disorder.
- Visualized distinct profiles using radar plots, enhancing the understanding of diverse patterns of suicidal thoughts.

EDUCATION

Columbia University | *Master of Art* ***2018 - 2020***

Clinical Psychology with a concentration in Research/Quantitative Methods (GPA: 4.0/4.0)

Coursework: Data Science; Machine Learning; Probability and Inferential Statistics; Applied Regression Analysis

Stanford University | *Exchange Program* ***2010***

National Taiwan University | *B.A. in Economics* ***2008 - 2012***