Hongling Yang | PhD, Data Scientist

Address: Temecula CA **Tel:** (915) 219-1391 **E-mail** hyang78227@gmail.com

LinkedIn (22) Hongling Yang | LinkedIn **Github** https://github.com/hyang78227

PhD-level data scientist with over 10 years of analytical experience. Proficient in statistical modeling, machine learning, and big data tools.Passionate about translating complex data into actionable business insights.

SKILLS

- Python (Numpy, Pandas, Scikit-learn, Matplotlib, Seaborn, SciPy, scikit learn, TensorFlow), SQL (MySQL, PostgreSQL), Jupyter, Tableau, Spark, Google Cloud, Linux, Jiar, Git, MS Project, Excel, Slack, SharePoint
- Machine Learning (Regression, Classification, Clustering), Optimization, Data Mining, Statistical/Predictive Modeling, Data Visualization, Logistic Regression, Decision Trees, Random Forest, GBoost, XGBoost, SVM, K-nearest Neighbor (KNN), Naïve Bayes, Feature Engineering, NLP (NLTK), Deep Learning (Tensorflow, Keras), Experiment Design & Analysis, A/B test

EXPERIENCE

Applied Data Scientist (Springboard Trained)

Sprinboard, San Francisco, CA 2023- 2023

- Developed a price optimization strategy for a Ski Resort using regression models (Multivariate Regression, Random Forest, Gradient Boost, XGBoost) and Scikit-Learn, streamlined through a Python data pipeline with hyperparameter tuning, resulting in a ~20% Increase in revenue.
- Designed a customer segmentation system for wine enthusiasts by applying K-means clustering and Principal Component Analysis (PCA) on 10,000+ data points, leveraging Scikit-Learn and Pandas. The approach, based on responses to wine offers, led yo a 15% increase in targeted marketing efficiency and a 10% rise in sales conversions.
- Leveraged ImageDataGenerator to enhance minority class data, and designed a convolution neural network (CNN) with Transfer, Learning achieving a 95%+ accuracy rate on fashion image classifications. Employed the Hyperband optimization for hyper-parameter adjustments and integrated the pre-trained VGG16 model for feature extraction, culminating in a robust item-based fashion recommendation system
- Executed an ARIMA time series analysis on Cowboy Cigarettes' (TM, est. 1890) historical data from 1949-1960, using pandas and statsmodels. Identified an upward sales trend with seasonal summertime boosts, forecasting a 15% annual increase and a notable 25% summertime sales surge.

Data Scientist

2016- 2017

School Of Medicine, University Of California, San Diego, CA

- Analyzed Uganda's Rakai District data on 5,000+ participants, linking alcohol use and intimate partner violence (IPV) to HIV trends.Piloted an intervention reducing alcohol and IPV among men by 15%, setting a foundation for broader implementation.
- Studied 3,000+ African American women in Baltimore, identifying a 20% increased HIV risk due to neighborhood factors and forced sex occurrences. Utilized multilevel analysis to highlight key stress pathways, aiding prevention efforts.

Statistical Consultant

2010- 2014

Texas Tech Health Center, El Paso , TX

- Guided medical research, mentored residents, and partnered with physicians in clinical trials.
- Examined 25 hydroxy vitamin D's impact on 43 hemodialysis patients in El Paso TX (2009-2011). Results indicated a 30% PTH level reduction and 20% improved calcium regulation, highlighting vitamin D's therapeutic potential in ESRD treatments.

Statistician & Lecturer

2008-- 2016

College of Engineering, University of Texas, El Paso, TX

- Conducted GIS research targeting pollution analysis and collaborated on various academic papers. Instructed courses in statistics and mathematics.
- Applied land-use regression (LUR) and PCA to refine intraurban air pollution modeling for PM2.5 across 13 monitoring sites. Developed a vehicle emissions surrogate with an enhanced predictive capacity (R^2= .458), improving LUR model accuracy and boosting its cost-effectiveness.
- Utilized a functional linear regression model to probe the relationship between PM2.5 peaks and mortality during 8 a.m traffic surgesDiscovering a 2.4% uptick in mortality risk for a 10μg/m^3 PM2.5 increase, the findings, underscore the urgent need for improved urban traffic management to curtail pollution-related health threats.

CERTIFICATION & LANGUAGES

- SAS Advanced, Base, Clinical Trials Programmer for SAS 9
- Fluent in Chinese and English

EDUCATION

Ph.D	Statistics	Arizona State University	Tempe, AZ	2005 - 2008
M.S	Statistics	University of Texas at El Paso	ElPaso, TX	2003 - 2005
M.A	Finance	Peking University	Beijing, China	1997 - 2003

PUBLICATIONS

For a comprehensive list of my publications, please visit □hongling yang - □Google Scholar