Hongling Yang, PHD

LinkedIn : (3) Hongling Yang | LinkedIn
Address : Murrieta CA **Phone** : (915) 219-1391 **Email** : hyang78227@gmail.com **GitHub** : hyang78227 (github.com)

PROFESSIONAL PROFILE

Data science professional with a PhD in statistics and more than 10 years' experience and a strong statistical and analytical background. Expertise in data mining, python, R, SAS, C++, machine learning, and statistics. Passionate about solving problems using data, and presenting insights to business audiences

TECHNICAL SKILLS

DBMS: MS SQL Server, MySQL, Postgres
Analytical Tools: SQL, Python, R, SAS.
Data Science: Data Wrangling, Data Visualization, Statistical Modeling, Predictive Analytics, Forecasting Analytics, Data Wrangling, Data Pre-processing, Data Visualization, Statistical Modeling
Machine Learning: TensorFlow, Pytorch, Pyspark, Natural Language Processing, Recommendation Systems, Neural Network, Time Series, Clustering, Dimension Reduction, Bagging and Ensemble Methods, Logistic Regression, SVM, Naive Bayes, Hadoop Ecosystem , Image Processing, Optimization, unstructured Data

RELEVANT EXPERIENCE

Career Track, Springboard 2023-2023

➢ Project: Fashion Product Image Classification with Convolution Neural Network

GitHub - hyang78227/CapstoneProjectTwo

- Data Augmentation: Enhanced minority class representation using ImageDataGenerator.
- Deep Learning: Achieved >95% classification accuracy with a CNN via Transfer Learning and Employed VGG16 for feature extraction and built a fashion product recommendation system.
- Optimization: Applied Hyperband algorithm for hyperparameter tuning

➢ Project: A Google App Store Educational Apps Rating Analysis

GitHub - hyang78227/capstone-project3

- Data Exploration: Conducted EDA following data wrangling, pre-processing, and visualization techniques
- Classification: Used Decision Tree, Random Forest, and Gradient Boosting classifiers for prediction of the rating class ('Low', 'High') of educational Apps
- Imbalance Resolution: Address class imbalance using imbalanced-learn module
- Optimiation: Applied Hyperband for efficient hyperparameter tuning of classifiers.

➢ Project: Big Mountain Ski Resort Ticket Pricing Study

GitHub - hyang78227/DataScienceG

- Modeling: Employed Multivariate Linear Regression and Random Forest Regression, achieving a $19/ticket value increase. Pipeline Design: Streamlined data processing, regression, tuning, and model selection using a structured pipeline, resulting in a cohesive Python notebook.
- Optimization: Leveraged GridSearchCV for precise hyperparameter tuning and model training.

➢ Other Projects

GitHub - hyang78227/Springboard

- COVID-19 Patient State Classifications: Analyzed South Korean COVID-19 data; Employed Random Forest Classifier to predict patient states: 'isolated', 'released', 'deceased'
- Flight Departure Delay Prediction: Utilized Light GBM to predict flight delays ≥15 minutes; Used Bayesian Optimization for hyperparameter tuning and applied feature engineering for enhanced feature extraction
- Cigarette Sales Time Series Analysis: Analyzed Cowboy Cigarettes' historical sales data; Conducted time series analysis to forecast sales trends.
- Wine Customer Segmentation: Executed K-means clustering on wine customer data; Segmented customers based on their wine offer responses

Statistician, School of Medicine, University of California, San Diego 2016-2017

- Rakai District Study: Analyzed associations between alcohol use and IPV; explored the alcohol-IPV-HIV infection relationship using structural equation modeling.
- Alcohol & IPV Intervention: Tested a 2-arm pilot intervention for men, estimating required effect sizes for broader application
- Baltimore HIV Risk Study: Used mixed methods to study how neighborhood factors affect forced sex rates among African American women, influencing HIV risk.
- Stress & HIV Risk Analysis: Employed a multilevel design to understand stress-related pathways between forced sex and HIV risk behaviors.

Statistical Consultant, Department of Internal Medicine, Texas Tech Health Center **2010-2014**
• Spearheaded medical research, mentoring residents throughout the process.
• Originated research concepts and outlined sampling strategies.
• Enhanced the resident program with expertise in Statistical Computing.
• Partnered with physicians in executing clinical trials.

Statistician, College of Engineering, University of Texas, El Paso **2008-2014**
Lecturer, Mathematical Sciences, University of Texas, El Paso **2008-2016**
• Geographic Information System (GIS) Research: Played an instrumental role in the 'Ride8 Project', targeting Ozone pollution analysis in El Paso, TX.
• Collaborative Research: Teamed up with fellow researchers, culminating in multiple published papers.
• Academic Endeavors: Imparted knowledge as an educator in mathematics and statistics courses.

EDUCATION

PhD	Statistics	Arizona State University	Tempe, AZ	**2005-2008**
M.S	Statistics	University of Texas, El Paso	El Paso, TX	**2003-2005**
B.S	Finance	Peking University	Beijing, China	**1998-2002**

OTHER CERTIFICATIONS

• SAS Certified Advanced Programmer for SAS 9 (Certified Serial Number: AP011585v9)
• SAS Certified Base Programmer for SAS 9 (Certified Serial Number: BP038502v9)
• SAS Certified Clinical Trials Programmer Using SAS 9 (Certified Serial Number: CTP001236v9

PUBLICATIONS

Yang H. L., Li, W.W. et al. (2014). Development of a Principal Component Regression Model for Predicting Ozone Exceedance, *Air and Waste Management Association Annual Conference and Exhibition*, *Long Beach, CA. June 24-27th 2012.*
URL: Development of a principal component regression model for predicting ozone exceedance (researchgate.net)

Valenzuela V., Yang H. L. et al. (2012). Conceptual Modeling of Ozone Pollution for an Air Quality Basin in Texas, *Air and Waste Management Association Annual Conference Proceedings*, *San Antonio, TX. 2012.*

Yang, H. L (2015). A Study of Additive Coefficient Models, *under review by Journal of Statistical Planning and Inference*
URL: A study of additive coefficient models - ProQuest

Alhamad,T., Blandon, P., Yang, H. L. (2011). Effectiveness of 25 Hydroxy Vitamin D on Secondary Hyperparathyroidism in Hemodialysis Patients. *June 2011 Issue of Renal And Urology News.*

Olvera, H., Li, W.W., Yang, H. L. (2011). PCA Optimization of $PM_{2.5}$ Land Use Regression Model with Small Monitoring Network.

Submitted *to Sci Total Environ. Submission No.: STOTEN-D-11-02680*

URL: https://pubmed.ncbi.nlm.nih.gov/22464030/

Staniswalis, J., Yang H.L. (2009). Using a Continuous Time Lag to Determine the Associations Between Ambient PM2.5 Hourly Levels and Daily Mortality: Indication of the Importance of the Total Number of the Particles, *Journal of Air & Waste Management*, 59:1173–1185, *DOI: 10.3155/1047-3289.59.10.1173*

A Further Study of the Relationship between PM10 Level and Daily Mortality in El Paso, Texas Using a Historical Functional Linear Model. *M.S. Thesis (2005).* University of Texas; El Paso: 2005. AAT 1430952.

URL: https://scholarworks.utep.edu/dissertations/AAI1430952/