Hongling Yang
Data Scientist

Address Temecula, CA 92593
Phone (915) 219-1391
E-mail hyang78227@gmail.com

LinkedIn www.linkedin.com/in/hongling-yang-909773227
Github https://github.com/hyang78227

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

---

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

---

EDUCATION

PhD: Statistics   2005 - 2008
Arizona State University - Tempe, AZ

M.S: Statistics University   2003 - 2005
Texas - El Paso El Paso, TX

B.S: Finance   1998 - 2002
Peking University - Beijing, China

EXPERIENCE

Data Science Trainee   2023-2023
Sprinboard, San Francisco, CA

➢ **Fashion Product Image Classification** - GitHub - hyang78227/capstone-project3
  • Achieved 97.5% accuracy with CNN via Transfer Learning; enhanced minority class representation and implemented Hyperband for optimization.
➢ **Google App Store Educational Apps Analysis** - GitHub - hyang78227/CapstoneProjectTwo
  • Predicted app ratings with up to 94% accuracy using multiple classifiers; addressed class imbalance and optimized using Hyperband.
➢ **Big Mountain Ski Resort Pricing Analysis** - GitHub - hyang78227/DataScienceG
  • Applied regressions to increase ticket value by up to $19; streamlined data pipeline and improved model accuracy to 96% with GridSearchCV.

Other Projects - GitHub - hyang78227/Springboard
  • **COVID-19 Analysis**: Classified patient states with 92% accuracy using a Random Forest Classifier on South Korean data.
  • **Flight Delay Prediction**: Predicted delays with 94% accuracy using Light GBM; refined precision via Bayesian Optimization.
  • **Cigarette Sales Forecast**: Analyzed historical data to project sales trends; achieved a MAPE of 3.5%.
  • **Wine Customer Segmentation**: Segmented wine customers with a silhouette score of 0.75 using K-means clustering.

Statistician
2016-2017

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

- **District Study**: Investigated alcohol use's correlation with IPV and its subsequent HIV risk using structural equation modeling.
- **IPV Intervention**: Piloted a 2-arm intervention for men, gauging effect sizes for wider deployment.
- **Baltimore HIV Study**: Assessed neighborhood influences on forced sex rates among African American women, delineating HIV risk.
- **Stress & HIV Analysis**: Utilized multilevel analysis to discern stress-driven pathways linking forced sex to HIV risk behaviors.

Statistical Consultant
2010- 2014

Texas Tech Health Center, El Paso , TX

- **Medical Research**: Led research initiatives, mentored residents, and formulated sampling strategies.
- **Program Enhancement**: Elevated the resident program with advanced statistical computing skills.
- **Clinical Trials**: Collaborated with physicians for seamless execution.

Statistician & Lecturer
2008-- 2016

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

- **GIS Research**: Pivotal in the 'Ride8 Project' for Ozone pollution analysis in El Paso, TX.
- **Research Collaboration**: Co-authored several published papers with research peers.
- **Teaching**: Instructed courses in mathematics and statistics.

CERTIFICATION

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)

LANGUAGES

Chinese, English