I-Huei (Melanie) **Ho**

Data Analyst · Data Scientist

4021 Dominion Cove, Austin TX, 78759

(512) 420-3825 | melanieihuei@gmail.com | melanieihuei.github.io | melanieihuei | ihuei-ho

Master of Science in Statistics at University of Georgia. Seeking a **data analytics** related position. Experienced in **statistical modeling**, **natural language processing**, and **image processing** with over 5 years experience in R and 2 years experience in Python and PySpark.

Education

M.S. in Statistics University of Georgia — *Athens, GA* — *Aug. 2016 - May 2018*

Courses: Data Science Practicum in Python and Spark, Data Science, Data Mining, Statistical Programming in R and SAS, Linear Model, Probability Theory, Statistical Inference, Statistical Consultation, Time Series, Sampling Method

B.S. in Statistics National Cheng Kung University — *Tainan, Taiwan* — *Sep. 2010 - Jun. 2014*

Data Analytics Experience

Web Traffic Forecasting of Wikipedia pages — *Athens, GA*

Dept. of Statistics and Dept. of Computer Science, University of Georgia — *Jan. 2018 - May 2018*

- Retooled R time series package `itsmr` into Python version `itsmpy`
- Modularized ARIMA and Long Short-term Memory (LSTM) models applied to 145k Wikipedia pages in Python and resulted in 38.89 mean symmetric mean absolute percent error

Microsoft Malware Classification on Apache Spark — *Athens, GA*

Dept. of Computer Science, University of Georgia — *Feb. 2018*

- Features mining from .bytes and .asm files and features reduction via inverse document frequency (IDF) value and decision trees
- Applied random forest classifier on Pyspark by submitting jobs to Google Cloud computing machine and resulted in 98.97% accuracy of malware classification

Automatic Emotional Detection from Image Data with Support Vector Machine — *Athens, GA*

Dept. of Computer Science, University of Georgia — *Sep. 2017 - Dec. 2017*

- Extracted histogram of oriented gradients feature descriptors of images using OpenCV in Python
- Optimized support vector machine for 20 minutes process time of 1300 images and resulted in 98.1% testing accuracy

Genetic Serious Disease Investigation — *Athens, GA*

Dept. of Statistics, University of Georgia — *May. 2017*

- Filled missing values of blood tests results through regression trees in R
- Graphed Assumption checks and diagnostics in ggplot2 and established logistic regression model and time series model in R

Other Work Experience

Graduate Teaching Assistant — *Athens, GA*

University of Georgia — *Aug. 2017 - May 2018*

- Provided lectures of implementing R in regression models for social datasets in course SOCI6630
- Held several workshops addressing application of R at Department of Sociology

Associate Analyst of Supply Chain Management Division — *New Taipei, Taiwan*

Everlight Electronics Co., Ltd. — *July 2014 - Aug. 2015*

- Assessed and predicted future stock depreciation for monthly skull session and resulted in 15% sales revenues increase and one plant extension in southern Taiwan
- Evaluated potential devaluated products, demonstrated price-reducing trend to sales management division, and prevented 60% possible depreciation

Surveying Analyst — *Tainan, Taiwan*

National Cheng Kung University — *Oct. 2011 - Jan. 2012*

- Assisted investigations with pool surveys of 2012 Taiwan Presidential election and Legislative Election by collaborating with industries
- Submitted reports for improvement of obtaining reliable results and increased survey completion rate by 15%

Skills

Programming	R, Python
STAT Models	Generalized Linear Model, Mixed Effect Model, Logistic Regression, ARIMA Model, ANOVA
ML Models	Logistic Regression, Support Vector Machine, k-NN, Random Forest Classifier, Constrained NMF, PCA, SVD
Data Science tools	NumPy, Pandas, Scikit-learn, Scikit-image, NLTK, OpenCV, TensorFlow
Data Visualization	ggplot2, Matplotlib, Tableau, Html
Other Tools	Apache Spark, MySQL, Unix, Git, SAS, Microsoft Excel, Google Computing Engine, LaTeX, R Markdown, Jupyter Notebooks