

# I-Huei (Melanie) Ho

DATA SCIENTIST

639 Geary St, San Francisco, CA 94102

☎ (512) 420-3825 | ✉ melanieihuei@gmail.com | 🏠 melanieihuei.github.io | 📷 melanieihuei | 🌐 ihuei-ho

Data Scientist at Holo, Inc., and Master of Science in Statistics at University of Georgia. Experienced in **statistical modeling**, **natural language processing**, and **computer vision** with over 3 years experience in Python and Pyspark, and 6 years experience in R.

## Education

<b>M.S. in Statistics</b>	UNIVERSITY OF GEORGIA	Athens, GA	Aug. 2016 - May 2018
<b>B.S. in Statistics</b>	NATIONAL CHENG KUNG UNIVERSITY	Tainan, Taiwan	Sep. 2010 - Jun. 2014

## Data Analytics Experience

### Data Scientist

HOLO, INC.

Oakland, CA

Jun. 2018 - Present

- Surface data in Airtable and maintain pipelines using Azure Data Factory and Azure Functions
- Programmatically label defects in cast films using OpenCV and ML clustering and evaluate feature importance through PCA for predicting resin usage
- Design and deploy Django app for company-wide data collection API hosted on Azure and create ETL jobs that consumed IoT messages and avro files using Azure web jobs and Databricks
- Build analytics dashboard and present real-time printing information with plotly dash, PowerBI and Domo

### Web Traffic Forecasting of Wikipedia pages 🔗

DEPT. OF STATISTICS AND DEPT. OF COMPUTER SCIENCE, UNIVERSITY OF GEORGIA

Athens, GA

Jan. 2018 - May 2018

- Retooled R time series package **itsmr** into Python version **itsmppy**
- 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 🔗

DEPT. OF COMPUTER SCIENCE, UNIVERSITY OF GEORGIA

Athens, GA

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

### Scalable Document Classification on Apache Spark

DEPT. OF COMPUTER SCIENCE, UNIVERSITY OF GEORGIA

Athens, GA

Jan. 2018 - Feb. 2018

- Created large-scale Naive Bayes document classifier based on word counts on Apache Spark and resulted in 94.52% accuracy for the largest testing dataset working on Google Cloud Platform
- Optimized Naive Bayes classifier by implementing punctuations and stop-words removing and words stemming, and Laplace smoothing to zero-counts words in each label class

## Other Work Experience

### Graduate Teaching Assistant

UNIVERSITY OF GEORGIA

Athens, GA

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

EVERLIGHT ELECTRONICS CO., LTD.

New Taipei, Taiwan

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

## Skills

<b>Programming</b>	Python, R
<b>STAT Models</b>	Generalized Linear Model, Mixed Effect Model, Logistic Regression, ARIMA Model, ANOVA
<b>ML Models</b>	Logistic Regression, Support Vector Machine, k-NN, Random Forest Classifier, Constrained NMF, PCA, SVD
<b>Data Science tools</b>	NumPy, Pandas, Scikit-learn, Scikit-image, NLTK, OpenCV, Keras, Tensorflow
<b>Data Visualization</b>	plotly, dash, ggplot2, Matplotlib, Tableau, Html
<b>Other Tools</b>	Django, PySpark, MySQL, Unix, Git, GCP, Azure, LaTeX, R Markdown, Jupyter Notebooks