

# ZHAOYUN MA

## Data Scientist

(412) 596-3997 • [mazhaoyunyh@gmail.com](mailto:mazhaoyunyh@gmail.com) • [LinkedIn](#) • [GitHub](#) • Columbia, SC

### SUMMARY

- Highly motivated data scientist with advanced analytics skills and ability to communicate complex concepts cross-functionally
- Strong technical, business, and problem-solving acumen. Content creator of “66 Days of data” on [GitHub](#), [Medium](#), [LinkedIn](#)
- US permanent resident, and fluent in Chinese and English

### EDUCATION

**Ph.D. in Mechanical Engineering**, *South Carolina, SC* 2017 – 2020  
University of South Carolina, Overall GPA: 4.0/4.0,  
**Master of Science in Civil Engineering**, *Pittsburgh, PA* 2015 – 2016  
University of Pittsburgh. Overall GPA: 3.9 /4.0

### PROFESSIONAL EXPERIENCE

**University of South Carolina**, *South Carolina, SC* 01/2021 – present  
**Research Associate**

- Collected, structured, analyzed, and visualized wavefield data of target structure in MATLAB
- Developed a filter-based network imaging algorithm to extract material thickness features and achieved 3  $\mu$ m accuracy
- Spearheaded the DOE material evaluation project to detect material degradation and prevent nuclear structural failure
- Organized/created a training program for programming, data processing and device operation

**Research Assistant** 01/2017 – 12/2020  
[[Conference Paper](#)]

- Automated wavefield data collection by developing a noncontact laser inspection system
- Invented a filter-based network imaging algorithm in MATLAB for complex damage profiling [[US patent](#)], [[Journal Paper](#)]
- Led the NASA advanced composite project at UofSC and cocreated a Nondestructive Evaluation Handbook [[Journal Paper](#)]
- Taught mechanical engineering lab course for sensor installation and data measurement

**Advanced Skills:** Research, Mechanical Engineering, Data/Signal Processing, Image Processing, Data Visualization, Statistics, Data Science, Machine Learning, Deep Learning

### PROJECTS

**Sentiment Analysis Web App Deployment (NLP, RNN)** [[GitHub](#)]

- Deployed an LSTM RNN model using PyTorch with test accuracy of 87% to a Web App through Amazon SageMaker and S3
- Extracted features using text preprocessing and Bag of Words, and trained the LSTM model with hidden dimension as 200

**Dog Breed Classifier (Computer vision, CNN)** [[GitHub](#)]

- Trained and deployed a dog breed classifier using transfer learning with test accuracy 87% through Amazon EC2
- Implemented a human face detector using OpenCV and a dog face detector using VGG16

**Customer Segmentation Report** [[GitHub](#)] [[Medium](#)]

- Preprocessed/structured the raw data and Identified core customer segments using PCA and K-Means
- Predicted customer conversion probability with 0.8 AUROC score using XGBoost and improved the model using Bayesian Optimization and GridSearchCV

**Boston Airbnb Listing Price Estimator** [[Kaggle](#)] [[Medium](#)]

- Trained data using different models (linear/ridge regression, neural network) to predict the listing price with 73% accuracy
- Created a heatmap that shows top 15 numerical and categorical features that have highest influence on the price

**Disaster Response Classification** [[GitHub](#)]

- Created a Web App to run a disaster classification model with provided input message and classify its disaster categories
- Built ETL pipeline to read, clean and store data into SQLite database and built a machine learning pipeline to train and export the best classifier within the selected hyperparameters

### SKILLS

**Programming:** Python, SQL, R, MATLAB  
**Modules & Libraries:** Pandas, Numpy, Matplotlib, Scipy, NLTK, sklearn, PyTorch, TensorFlow, Keras, ggplot, dplyr, tidyverse  
**Data Analytics:** Data Mining and Visualization, Quantitative Analysis, ETL pipeline  
**Cloud computing:** Amazon SageMaker, Amazon S3, Amazon EC2, SQLite  
**ML/AI, Deep Learning:** Predictive Modeling, Linear Models, Boosted Tree Models, Machine Learning Pipeline, Model Deployment, Recommender Systems, Convolutional Neural Network, Recurrent Neural Network, Generative Adversarial Network, Computer Vision, Natural Language Processing

### CERTIFICATIONS & ACTIVITIES

**Certifications:** [Data Scientist Nanodegree](#) (Udacity, 2020), [Machine Learning Engineer Nanodegree](#) (Udacity, 2021), [Deep Learning Nanodegree](#) (Udacity, 2021), [R Programming](#) (Coursera)

**Social Impact:** Mentor of 3 high school student interns, Research Reviewer of Discover at UofSC Program