# MEEYOON CHOO

2495 Sawdust Rd, Apt 705, Spring, TX 77380.
(929)409-9786 ⋄ meeyoonchoo@yahoo.com ⋄ http://meeyoonchoo.github.io

## EDUCATION

**Master of Science in Applied Data Science** *August 2018 - December 2019*
Bay Path University, Longmeadow, MA
**Master of Science in Biological Science** *January 2010 - August 2012*
The University of Memphis, Memphis, TN
**Bachelor of Science in Biotechnology** *May 2013 - May 2016*
The University of Malaysia Sarawak, Sarawak, Malaysia

## COURSE PROJECTS

**Prediction of Heart Disease Using Machine Learning Techniques From Kaggle Challenge**
The project goal's to train the data with variety machine learning techniques and investigate them in order to find the best model to predict the presence of heart disease. ML techniques used such as Logistic Regression, K-Nearest Neighbor (KNN), Principle Component Analysis (PCA), Random Forest, and Support Vector Machine (SVM). Compared and evaluated all the models with confusion matrices and accuracy rates. The data set is from https://www.kaggle.com/ronitf/heart-disease-uci.

**Data Exploration and Visualization**
Used PROC SQL in SAS to sort, modify, retrieve, and summarize the data set about Ebola virus disease outbreak in 10 affected countries in West Africa. Additionally, TIBCO Spotfire was used to analyse and build data visualization on a dashboard. The data set comes from The Humanitarian Data Exchange (HDX) (https://data.humdata.org/dataset/ebola-cases-2014).

**Database Management System & Design**
Designed movie theater system to help users to keep the records of all movies, maintain or update the records of movies, easy access to database containing information, and calculate the total amount of tickets sold. It also allowed the users to handle the work efficiently and gain customer satisfaction and help to increase reliability.

**Data Forecasting of Time Series**
Used various benchmark methods (Drift, Mean, Naive, Snaive Method) in R to forecast for sales of new-one family houses.

## SKILLS

Statistical Modeling and Analysis, Multiple Linear and Logistic Regression, QDA, KNN, SVM, PCA, Decision Tree, Random Forests
Software, Languages & Tools: SAS, R, Python, SQL, TIBCO Spotfire, MS Office

## RELATED WORK EXPERIENCE

**The University of Memphis. Position: Research Assistant** *January 2010 - August 2012*
Collaborated with USGS and National Park Service (Canyonlands National Park), studied the spatial level of relationships between native grass species and geographical features that shape the patterns of genetic diversity across the landscape using statistical tools such as AMOVA, pairwise Fst, Principal Coordinate Analysis (PCA), Isolation by Distance (IBD), etc. Ultimately, developed guidelines and strategies for restoration of degraded grassland in Canyonlands National Parks.