Nguyen Nguyen

nnguyenlethanh@mbs.miami.edu

Education

UNIVERSITY OF MIAMI   Miami, FL
PhD Marketing   Expected May 2025

UNIVERSITY OF MIAMI   Miami, FL
B.S Computer Science, minor Mathematics.   May 2020
GPA: 3.967/4.0
GRE: 167/170, 90th Percentile, Quantitative

Relevant Coursework: CS224W Machine Learning with Graph, CS224N Natural Language Processing, Big Data Mining, Database Systems, Applied Econometrics, Intro to Probability & Statistic, Linear Algebra, Python for Data Scientist, Data Structure & Algorithm Analysis, Theory of Computing, Intro to Marketing
Certifications: IBM Professional Certification Data Scientist, Stanford Deep Learning Coursera, Udacity Deep Learning, Udacity Intro to Machine Learning

Publication

- Nguyen, Nguyen Le Thanh, et al. "Multi-Scale Auralization for Multimedia Analytical Feature Interaction." *Audio Engineering Society Convention 147*. Audio Engineering Society, 2019.
- Nguyen, Nguyen Le Thanh, et al. "A new auditory image for social media: Moving towards correlation of spectrographic analysis and interpretation with audience perception." *The Journal of the Acoustical Society of America* 146.4 (2019): 2846-2846.

Projects

- CS224W: Machine Learning with Graphs
    - Implement from scratch GraphSAGE, GAT (Graph Attention) for homogeneous graph use PyTorch Geometric
    - Implement HAN (Wang et al, 2019) for heterogenous graphs
    - Learn and use NetworkX to construct Graph datasets
    - Learn and use DeepSnap to split Graph datasets for tasks such as link prediction, node prediction, graph prediction
    - Github: https://github.com/hdvvip/CS224W_Winter2021

- AWS SageMaker Sentimental Analysis Web App:
    - Trained and Deployed XGBosst model on Amazon AWS to perform sentimental analysis on movie reviews
    - Preprocessing 25,000 movie reviews by using techniques such as drop stop words, tokenization
    - Deploy and train XGBoost model to predict whether the movie review is positive or negative
    - Create an Endpoint, Lambda function, and Amazon API Gateway to connect the deployed model on AWS to a web app
    - Built a simple web app which interacts with the XGBoost model performing sentiment analysis on movie reviews
    - Github: https://github.com/hdvvip/LSTM-Sentimental-Analysis-Sagemaker-Deployed

- Seinfeld TV Show Script Generator:
    - Preprocessing the script of 9 seasons of Seinfeld TV show
    - Create a Lookup Table to tokenize all text data to numeric vectors
    - Create Pytorch Data Loader with training data sample of 10 continuous words and the target variable is the 11th word in the sequence
    - Train an RNN Sequence-to-Sequence model to generate a novel Seinfeld TV script
    - Github: https://github.com/hdvvip/Seinfield-TV-Show-Script-Generator

- AI Image Classifier:
    - Allowing users to train and save the new model on a user customize dataset of images and then predict their classes from the command line.
    - Allowing users to choose a set of pre-trained ImageNet Deep Learning model from Pytorch Torchvision Models: AlexNet, VGG, ResNet, GoogLeNet, Inception, etc
    - Automatically detect and train the model on GPU if the user has one.
    - Github: https://github.com/hdvvip/AI_Image_Classifier

- Customer Clustering: AZDIAS Germany Insurance Company

    - Processed data, including handling missing values, discretizing continuous variables, features scaling for more than 3 million rows and 700 columns of Germany demographic data.
    - Used Principal Component Analysis for dimensionality reduction, reduce original data of 700 columns to 109 columns
    - Use K-Means clustering to cluster customers into groups based on their demographic data
    - Achieve 0.035 Silhouette Score with K=10
    - Reverse transform PCA to find out which group of customers has the highest potential making the insurance purchase
    - Github: https://github.com/hdvvip/Identify_Customer_Segments/blob/master/Identify_Customer_Segments.ipynb

Skills
Programming & Machine Learning Library: Python, R, MATLAB, PyTorch, TensorFlow, Keras, sklearn
Language: Vietnamese, Chinese