VAN MINH NGUYEN

(321)-368-9310 ◇ vmnguyen251@gmail.com
Melbourne, FL USA
LinkedIn Link, GitHub Link

CAREER OBJECTIVE

Highly motivated, exceptionally fast learner Graduate student in Operations Research looking to advance into a career in Data Sciences/Analytics. Aiming to utilize my academic and internship experience to breach the wall between theoretical mathematics (especially in Topological Data Analysis) and real-world applications.

EDUCATION

Florida Institute of Technology, Melbourne FL

Ph.D Operations Research GPA: 4.00 *Aug 2020 - Present*
M.S. Operations Research GPA: 4.00 *Aug 2018 - May 2020*
B.S. Biochemistry (Biology Emphasis) GPA: 3.43 *Aug 2014 - May 2018*

EXPERIENCE

Florida Institute of Technology - Dept. of Mathematical Sciences *August 2018 - Present*
Graduate Student/Teaching Assistant

- Researched on Double Stochastic Processes (Branching Process with random offsprings)
- Researching Electronics Health Record embeddings for Machine Learning/Deep Learning
- Teached and grade exams for Calculus I, II, III
- Tutored and grade exams for Probability & Statistics, Neural Networks.
- Aided students studying Stochastic Modeling, Differential Equations

Truveta *May 2021 - August 2021*
Graduate Intern

- Developed a data quality measurement toolkit, with **Databrick** and **PySpark**, to measure data quality of Truveta Health Data Model (THDM), garnering trusts from health providers.
- Designed a Monte Carlo sampling model to generate synthetic patient data for stress-testing.
- Established the foundation approach for (new) Synthetic Patient Health Data, involving probabilistic theory of document retrieval, representation (feature) learning and deep learning.
- Developed an annotation recommender system for THDM medical concept normalization.
- Assisted team members and other co-workers on various Projects and Collaborative efforts.

IBM Skill Academy *May 2021 - June 2021*
IBM Data Science Practitioner - Instructor

- Earned Certification for Instructor role for IBM Data Science Practitioner
- Audited IBM Data Science learning material and proposed changes that was added to the material.
- Created a curriculum for teaching IBM Data Science course at Florida Institute of Technology

PROJECTS

Classification of Extended MNIST Handwritten Digits and Letters Dataset
Project link

- EMNIST dataset is MNIST (handwritten digit) dataset with handwritten characters
- Applied Principal Component Analysis (PCA) and Persistent Homology (TDA) to reduce the dimension of dataset. Reduced feature size from 784 (28x28) to 35 while retaining 99% variance
- Utilized libraries **giotto-ai** along-side standard deep learning libraries **sklearn, NumPy, Tensorflow**
- Achieved 97%-91% training-testing accuracy with a tuned network with only 3 hidden layers

Google Speech Command Dataset Classification
Project link

- Classify Google Speech Command using Convolutional Neural Network on audio data (CNN)
- Pipeline processing audio data to image features with data augmentation
- Used multiple CNN architectures (LeNet, MiniGoogleNet, AlexNet) for training and stacking model for ensemble.
- Achieved 91% validation accuracy with stacking model for ensemble.

Sentiment Analysis on MyAnimeList User Ratings
Repo link, Project link

- Predict user rating based on review using Recurrent Neural Network (RNN)
- Setup a data-mining pipeline utilizing self-hosted REST API with a **Redis** server for caching inside **dockerized** container
- Used different models (RNN with LSTM, CNN, CNN with Word2Vec embedding layers) for training and stacking model for ensemble.
- Achieved 94% validation accuracy with ensemble model

Movie Ratings Analysis

- Initiate a Jupyter-Python notebook environment using **Amazon AWS EC Instance**
- Pre-processed/cleaned movie rating data extracted from IMDB.
- Import data to **SQL** database hosted on **Amazon AWS RDS** and ran queries to analyze user trends.

Date A Live: Spirit Pledge Assets Decryption tool
Personal Project. Repo link

- Reverse Engineer mobile game Date A Live: Spirit Pledge using Static analysis tool from NSA **ghidra** and dynamic analysis tool **frida**.
- Re-implement decryption functions using Python, implement methods to convert PowerVR, Ericsson Texture Compression format to digital images format (JPEG/PNG)
- Data-mined to find insecure API/server that allows easy download/extraction of new game contents.

Other Minor Projects
Personal Project. GitHub link

- Reverse engineer mobile game Arknights to find AES256 key for assets decryption. Repo link
- Maintained a toolkit that patches (cheats) mobile game Azur Lane. Added both 32/64-bit capability to the tool, features that allows user to obtain ingame paywalled contents for free while avoid cheating detection. Repo link
- Date A Live: Spirit Pledge Datamining repository using developed decryption tool. Using **cronjob** and **Github Action CI/CD** to automate fetch, decrypt and datamine new contents. Repo link

SKILLS

Software & Tools	Python, Spark/PySpark, MySQL, Redis, Tensorflow, PyTorch, ONNX
Operating Systems	Windows, Linux, UNIX-based (MacOS)
Skills	Data Processing & Analysis, Stochastic Modeling, Reverse Engineering
Languages	Vietnamese (Native), English (Fluent), Japanese (Basic)

ACADEMIC HONORS

- Dean's List for Spring 2016, Fall 2017, Spring 2018 (GPA $\geq$ 3.7)
- Member of **Phi Kappa Phi** Honor Society (top 10% percent of graduate students on campus)

PERSONAL TRAITS

- Enjoys adopting cutting edge technologies, especially in Deep Learning and Automation.
- Passionate and innovative researcher.
- Highly motivated, fast learner with exceptional multi-tasking skills.
- Love clearing up jargons with easy-to-understand concepts.