# Van Minh **Nguyen**

✉ vmnguyen251@gmail.com | 📱 (321)-368-9310 | 📍 Melbourne, FL USA
❧ Portfolio | ❧ LinkedIn Profile

## Education

**Florida Tech** — *Melbourne, FL*
PH.D OPERATIONS RESEARCH — *Aug 2020 - Present*
- Researched on Double Stochastic Processes (Branching Process with random offsprings).
- Researching Electronics Health Record embeddings for Machine Learning/Deep Learning.

**Florida Tech** — *Melbourne, FL*
M.S. OPERATIONS RESEARCH, GPA: 4.0 — *Aug 2018 - May 2020*

**Florida Tech** — *Melbourne, FL*
B.S. BIOCHEMISTRY (BIOLOGY EMPHASIS), GPA: 3.43 — *Aug 2014 - May 2018*

## Work Experience

**Truveta** — *Seattle, WA*
GRADUATE INTERN — *May 2021 - Present*
- Developed a toolkit to measure data quality of Truveta Health Data Model (THDM), garnering trusts from health providers. Contributed to new partnerships with multiple after Private Review in Aug 2021, and with Microsoft in Sept 2021.
- Established the foundation approach and model for Synthetic Patient Health Data, helps identify bottleneck in ETL process of EHR.
- Developed an annotation recommender system for medical concept normalization. Reduced internal annotators' workload by 80%
- Deployed cost-effective and scalable SOTA Named Entity Recognition (NER) Models for clinical notes information extraction. Saved $2 million in cost compared to existing third party solutions, and optimized to reduce 75% operating cost overhead compared with original baseline

**Florida Tech - Dept. of Mathematical Sciences** — *Melbourne, FL*
TEACHING ASSISTANT — *August 2018 - December 2021*
- Teached and grade exams for Probability & Statistics, Neural Networks, Calculus I, II, III.
- Aided students studying Stochastic Modeling, Differential Equations.

**IBM Skill Academy**
IBM DATA SCIENCE PRACTITIONER - INSTRUCTOR — *May 2021 - June 2021*
- 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 Tech.

## Skills

**Software & Tools:** Python, Spark/PySpark, MySQL, Redis, Tensorflow, PyTorch, ONNX
**Operating Systems:** Windows, Ubuntu Server, Arch Linux, macOS
**Cloud Infrastructure:** Microsoft Azure, Amazon AWS
**Other Skills:** Data Processing & Analysis, Stochastic Modeling, Reverse Engineering, Penetration Testing

## Projects

**Classification of Extended MNIST Handwritten Digits and Letters Dataset** — *giotto-ai, sklearn, Tensorflow*
APPLY PERSISTENT HOMOLOGY (TDA) FOR FEATURE ENGINEERING. 91% TESTING ACCURACY WITH SHALLOW NET. — *[Project link]*

**Sentiment Analysis on MyAnimeList User Ratings** — *Redis, Docker, Tensorflow*
DATAMINE, ETL TO EXTRACT USER REVIEW FROM MAL. PREDICT USER RATING BASED ON DATA-MINED REVIEW TEXT WITH SENTIMENT ANALYSIS. 94% TESTING ACCURACY WITH ENSEMBLE MODEL — *[Project link]*

**GPU PySpark Notebook with DeltaLake** — *Docker, Compose, Github CD/CI*
DATA ANALYSIS CONTAINER WITH JUPYTER NOTEBOOK SERVER WITH GPU-RAPIDS AI FOR GPU-ACCELERATED DATA TRANSFORMATION, SPARK/PYSPARK FOR DISTRIBUTED AND SCALABLE ETL, WITH ADDITION OF ALL COMMON DATA SCIENCE LIBRARIES. AIMING FOR FEATURE PARITY WITH DATABRICKS WHILE STAYING MINIMAL. — *[Repo link]*

**Date A Live: Spirit Pledge Assets Decryption tool** — *frida, ghidra, Github CD/CI*
REVERSE ENGINEERED GAME FOR ASSETS DECRYPTION. CRON PIPELINE FOR CONTINUOUS UPDATE AND DECRYPT DATA. — *[Repo link],[Data link]*