MY (EMMY) PHUNG

emmy.phung@nyu.edu
emmyphung.github.io
linkedin.com/in/emmy-phung
(860) 796-8237

EDUCATION

NEW YORK UNIVERSITY, *M.S. in Data Science* | **GPA: 4.00/4.00** — New York, NY
- Coursework: Deep Learning, Machine Learning, Big Data, Probabilistic Time Series, Probability & Statistics — May 2021
- **Grace Hopper Conference 2020 Scholar**
- 3rd Prize Winner, 2019 NYC Emerging Case Competition by Google, EY, PwC

TRINITY COLLEGE, *B.S. in Economics* | **GPA: 3.99/4.00** (Top 5%, Summa Cum Laude) — Hartford, CT
GOOGLE Summer 2019 South East Asia Machine Learning School — Indonesia

TECHNICAL SKILLS

- Programming/ Querying: **Python, PyTorch, PySpark, SQL, MATLAB**
- Cloud Computing: **AWS** (SageMaker, S3, EMR)**, Hadoop, Spark** | Analytics Software: **Tableau, Stata, Excel**
- **Machine Learning** (Linear & Logistic Regression, Decision Tree, Random Forest, XGBoost)**, Natural Language Processing** (BERT), **Time Series Modeling** (LSTM), **Recommender System** (ALS), **Deep Learning** (RNN, GRU), **A/B Testing**

PROFESSIONAL EXPERIENCE

Fidelity Investments, Inc. — Boston, MA
Data Science & Optimization Intern — June – August 2020
- Spearheaded the team's **first-ever** attempt in simulating volatility of S&P 500 and other fixed income assets in long time horizon (10+ years) using **LSTM with auto-encoder**; trained and deployed the model on **AWS Sagemaker**
- Enhanced data quality and **automated existing data-cleansing process** in Excel by creating a 3-step data imputation model that adopted Multivariate Imputation with Chained Equation (MICE) framework in Python
- Explored and implemented **new methods for back-testing** (ACF, PACF, GARCH, DTW, etc.); visualized future simulated assets in both 2-D & 3-D plots and video format (Plotly); presented findings to senior executives and portfolio managers

Comcast NBCUniversal & Techstars (3-month fulltime project) — Philadelphia, PA
Analytics Consultant — July – October 2018
- Conducted research on product-market fit and user experience for 5 newly launched products and services
- For Tally TV, Inc.: proposed and performed A/B tests on **2 new mobile app features** to engage users, defined new engagement metrics, mapped out conversion funnel & tracking methods; resulted in **11.7% increase in user engagement**
- For Orai, Inc.: utilized unit-economics metrics to develop market sizing models and finalized business pitches to present on Demo Day; helped Orai **successfully raise $2.3M** in capital

Road2College — Philadelphia, PA
Data Analyst — November 2018 – August 2019
- **Initiated and built a predictive model** that estimates the amount of scholarship awards given to freshmen depending on test scores, financial, and admission data (MSE: 0.132) to support college admission consulting service
- Scrubbed and synthesized licensed government data of 8,000+ US colleges (2016 – 2018)
- Performed EDA to find patterns and detect unusual changes in school admission, published findings on the company's website with **Tableau interactive charts/stories** and **infographics**

PROJECTS (see more at emmyphung.github.io)

IBM-NYU Capstone: Training Generalizable End-to-End Speech-to-Intent Model (On-going)
- Apply transfer learning to build a generalizable **speech recognition model for intent detection** in data-scarce domains
- Perform data augmentation; pretrain model on 20 hours of speech data from various domains (airline, smart-home devices, etc.)

Predicting Psychiatric Readmission Rate Based on Clinical Notes (NLP)
- Predicted probability of readmission based on discharge notes using **pre-trained Clinical BERT embeddings** (AUC: 0.747)
- Visualized medical insights from electronic health records of 7,050 psychiatric patients (MIMIC III) with **BertViz**

Goodreads Book Recommender System in Spark
- Built an **ALS recommendation model in Spark** training on 2M+ user-item interactions (review, ratings, etc.)
- Manipulated **big data in Hadoop** and illustrated how items are distributed in the learned space using UMAP and t-SNE

How Can Restaurants Improve Their Yelp Profiles for Success?
- **Forecasted success rate** of a restaurant using both linear and tree-based models with hyper parameter tuning (AUC of 0.781)
- Extracted features from restaurants' Yelp profile (e.g. pet-allowed); identified key success attributes based on feature importance

LEADERSHIP

Data Science in Brief, *Founder & Executive Director* — U.S. – Vietnam (April 2019 – Present)
- **Connect, inspire, and guide young Vietnamese data enthusiasts** through their learning and career in data science
- Host **annual educational and career workshops** with guest speakers from Apple, Microsoft, Salesforce, Airbnb, Etsy, etc.