

# Tung Thanh Le

Website: <http://ttungl.github.io/>

Mobile Phone: 612-490-3605

U.S. Permanent Residency

Personal Email: [ttungl@gmail.com](mailto:ttungl@gmail.com)

## ❖ Education

- **University of Louisiana at Lafayette, USA**  
*Doctor of Philosophy (Ph.D.) in Computer Science*  
08/2013 – 12/2018
- **Kumoh National Institute of Technology, South Korea**  
*Master of Engineering (M.Eng.) in IT Convergence Engineering*  
09/2011 – 08/2013
- **University of Louisiana at Lafayette, USA**  
*Master of Science (M.Sc.) in Computer Science*  
08/2013 – 12/2016
- **Danang University of Technology, Vietnam**  
*Bachelor of Engineering (B.Eng.) in Electrical Engineering*  
08/2002 – 08/2007

## ❖ Professional Work Experience

### ▪ Senior Data Scientist

**NBCUniversal**

12/2021 – Present

- **Lift Measurements**: Responsible for building ETL data pipelines with Python, PySpark, R, SQL on **Databricks** and **SnowPark** for data processing, feature engineering, feature selection, using matching methods to create a balanced covariate distribution in control and treatment groups of observational data for lift measurements.
- **Multi-Touch Attribution Models**: Responsible for building the end-to-end MTA and Optimization models from data collection, data engineering, to implement the models, to determine the impact of advertising channels in terms of their contributions to the conversions for customer journey. Using Rule-based, Markov, Shapley-value approaches. Applied Budget Optimization Modeling using **Gekko** for MTA outcomes to maximize ROI for advertisers.
- **Face Recognition**: Responsible for building the celebrities recognition model using MTCNN, FaceNet, and SVM. MTCNN is used to capture facial areas from inputs. Faces captured are used for training FaceNet. SVM is used to classify new faces based on Face Embedding from trained FaceNet model. Implemented on **AWS EC2 cluster**.

### ▪ Data Scientist

**J.D. POWER**

07/2018 – 12/2021

- **Days-to-turn on Vehicles Prediction**: Responsible for building predictive models using data analytics, machine learning to predict days-to-turn target which determines how long it takes to sell a specific new car in the inventory. Implemented on **AWS** and **databricks** using Python and SQL, and **Tableau**.
- **PIN Transformation**: Building ETL big data pipelines from SAS to Python using BigQuery, PySpark, Python, Javascript for production on **AWS**, **GCP** platforms.
- **Online Social Review Analytics**: Responsible for building the reviews sentiment analysis using natural language processing (**NLP**) techniques such as text cleaning, feature engineering using outlier remover, lemmatization, N-grams tokenization; Utilizing **AWS Comprehend**, **SageMaker**, **Google Cloud NLP**.

### • Research Intern

Hanwha Thales, South Korea

08/30/2012 – 12/31/2012: Responsible for optimizing the network topologies for ships' built-in-network communication.

### ▪ Software Engineer

Unilab-DUT (Novas Technologies Ltd.), Vietnam

04/01/2008 – 06/01/2011: Responsible for software-hardware development.

### ▪ Summer Intern

Orion Technologies Co., South Korea

06/01/2012 – 08/30/2012: Responsible for programming network communication in ships.

### ▪ PCB Layout & Design Engineer

Acronics Systems, Inc – San Jose, CA (Vietnam office)

06/01/2007 – 03/30/2008: Responsible for designing PCB on high-speed circuit boards.

## ❖ Projects

- **Donation Analytics (Insight Data Engineering Challenge)**: As a data engineer working for political consultants whose clients are cash-strapped political candidates, they've asked for help analyzing loyalty trends in campaign contributions, namely identifying areas of repeat donors and calculating how much they're spending. Identify areas (zip codes) that could be sources of repeat campaign contributions.
- **Behavioral Cloning (Deep Learning)**: Built and trained a convolutional neural network to drive the car itself autonomously in a simulator using Tensorflow (backend) and Keras. Experimented with a modified Nvidia architecture. Performed image processing with brightness, shadow augmentation, and flipped images using OpenCV. Used dropout and Adam optimizer to generalize the network for driving multiple tracks. The datasets are used via Udacity's source for training the model. Trained the model on Amazon **AWS EC2** platform with GPU instances.
- **Creating Customer Segments**: Evaluated what types of customers, wholesale distributors have to help them make better, more informed business decisions on the changes of their customers. Used unsupervised learning techniques (K-Means Clustering) to observe any similarities exist between customers.
- **Network-on-Chip Optimization**: Designed the mathematical modeling for optimizing interconnections and energy efficiency in network-on-chip. Used **CPLEX**, **Gurobi** solvers, Python (**pyomo**), Matlab (heuristic algorithms), and machine learning algorithms for solving this optimization problem.

## ❖ Professional Certificates

- Certification of Machine Learning (2017)  
*Online Course – Stanford University*
- Certification of Natural Language Processing Specialization (2021)  
*Online Course – DeepLearning.AI*
- Certification of Statistical Learning (2018)  
*Online Course – Stanford University*
- Data Analytics Methods for Marketing (2022)  
*Online Course - Meta*

## ❖ Honors & Awards

- Graduate Teaching Assistantship, 09/2015 – 06/2018
- NSF Graduate Research Fellowship, 09/2013 – 08/2015
- Best Paper Award - 14th Conference on Electronics & Info. Communications 2012
- NIPA scholarship and NRF scholarship, South Korea, 09/2011 – 06/2013
- Samsung Thales scholarship for student travel in 12/2012
- Excellent student, Danang University of Technology, 2004 – 2007
- One of four honor students achieving highest score on graduation thesis (4/500) in 2007

## ❖ Computer Skills

- **Programming languages**: Python, Java, PySpark, Scala, BigQuery, Javascripts, SQL, C/C++, R, MATLAB, CPLEX/AMPL.
- **Frameworks/Libraries**: Databricks, Airflow, Tensorflow, Keras, Apache Spark, Snowflake, Snowpark, MLLib, Node.js, OpenCV, Scikit learn, PyTorch, Spacy, nltk, OpenAI, AWS products, H2O.ai and driverless AI platform, Trax by Google.
- **Data Visualization**: Tableau, Power BI.
- **Cloud Services**: Amazon AWS, Google Cloud Platform.