Tung Thanh Le

Website: http://ttungl.github.io/ Mobile Phone: 612-490-3605
U.S. Permanent Residency Personal Email: 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

❖ ***Interest***

- Algorithmic Optimization, Mathematical Modeling
- Machine Learning, Big Data, Deep Learning and Artificial Intelligence

❖ ***Professional Work Experience***

- **Lead Machine Learning Engineer**
  Thomson Reuters
  5/2023 – Present
- **Senior Data Scientist**
  NBCUniversal *12/2021 – 4/2023*
  - ***Lift Measurements:*** The goal is to measure the impact of advertising campaigns. Responsible for building ETL data pipelines with Python, PySpark, SQL on **Databricks** and **SnowPark** for data processing, feature engineering, feature selection, using matching methods such as propensity score matching for measuring the impact.
  - ***Multi-Touch Attribution Models****:* The goal is to determine which channels have high contribution to the conversion of the viewers/customers, therefore advising advertisers to reallocate their investment to maximize their return on investment (ROI) via optimization modeling. Responsible for building the end-to-end MTA and Optimization models to determine the impact of advertising channels in terms of their contributions to the conversions for customer journey.
  - ***Face Recognition:*** The goal is to help data labeling on celebrity faces/brand objects in advertising video clips for conducting analysis on who contributed high sales/conversion rates in the advertising campaigns. Responsible for building the 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 Deep Learning instance**.
- **Data Scientist**
  J.D. POWER *07/2018 – 12/2021*
  - ***Days-to-turn on Vehicles Prediction****:* The goal is to help the OEM/dealers planning to optimally re-stock their sales inventories based on days-to-turn 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** and **Streamlit** for dashboards.
  - ***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:** The goal is to help evaluating the in-store performance rating based on the customers' reviews of the banks across U.S. 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. Used Udacity's dataset for training model. Trained the model on Amazon **AWS EC2**.
- **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***

- Generative AI with Large Language Models (2023)
  Online Course – DeepLearning.AI
- Certification of Natural Language Processing Specialization (2021)
  Online Course *– DeepLearning.AI*
- Certification of Machine Learning (2017)
  Online Course – Stanford University
- Certification of Statistical Learning (2018)
  Online Course – Stanford University

❖ ***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.