





# Tan TRAN

## Data Scientist

 (084) 385080690

 tanthm1102@gmail.com

 /in/tanthm

 tanthml

## Education

### MSc., Computer Science

Information and communications technology

Specialization: Data Science

John von Neumann Institute

2015 - 2019 | Ho Chi Minh, Vietnam

### BEng., Computer Network & Communication

University of Information Technology

2009 - 2014 | Ho Chi Minh, Vietnam

## Technical Skills

### Programming

JavaScript • SQL •  $\text{\LaTeX}$

Python • R • Scala

## Experience

### June 2019 - Present Visiting Scholar

ATMRI - Nanyang Technological University - Singapore

- Focus on developing mathematical models, machine learning models, simulation engine, real-time data pipeline.
- Projects: Conflict Detection, Conflict Resolution and Scenario Generators

**Tools:** Python, Plotly, Redis, Docker, Jupyter Notebook, GitHub

### July 2016 - June 2019 Data Scientist

Tenpoint7 Inc. - Ho Chi Minh city - Vietnam

- Deliver more than 5 analytic features for addy.ai - an automatic social text mining platform.
- Deliver analytic and visualization reports to customers in several consultant projects.
- Contribute in text-data analytic pipeline include data reprocessing at large scale, topic modeling, entity recognition, sentiment and emotion extraction based on state of the arts machine learning models.
- Work closely with Data Engineering team, to deploy machine learning models to production system.
- Projects: addy.ai, Social Reviews Mining, Data Crawlers, Audit and Evaluation Predictive Models.

**Tools:** Python, R, Scala, spaCy, gensim, scikit-learn, Docker, Amazon Web Services (S3, EC2, EMR, Comprehend, SageMaker).

## Research

2018

### MSc. Candidate

John von Neumann Institute - Ho Chi Minh city - Vietnam

**Thesis:** "Trajectory Clustering in Air Traffic Management with Enriched Flight Phrase Identification"

- Propose a feature engineering method which transform flight trajectory to "bag of words" features. The transformed features can performs compatibly with Locality Sensitive Hashing to solve the trajectory clustering problems.
- Prepare a ADS-B dataset of flights to Changi Air-port over 3 months in order to validate propose method.

**Tools:** Python, scikit-learn, pandas

## Publications

Tran, T., Pham, D.T., Duong, Q. and Mai, A., "An Adaptive Hash-Based Text Deduplication for ADS-B Data-Dependent Trajectory Clustering Problem," The 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, 2019.

Duong, Q., Tran, T., Pham, D.T. and Mai, A., "A Simplified Framework for Air Route Clustering Based on ADS-B Data," The 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, 2019.