# QUAN DUONG

## Data Scientist

📞 +84929221419    @ quanduongbao@gmail.com    🔗 www.linkedin.com/in/quandb

## EXPERIENCE

### Data Scientist

**Tenpoint7**

📅 05/2017 - Ongoing    📍 Ho Chi Minh

- Build analytic solution that solve variety of use case for customers
- Develop company internal product by productionize NLP techniques
- Work closely with Data Engineering and Full Stack Development team to deploy Machine Learning model on the cloud

### Data Science Apprentice

**Tenpoint7**

📅 09/2016 - 05/2017    📍 Ho chi Minh

- R&D on Data Science filed such as Predictive Modeling, Text Mining, Features Engineering, Clustering, Visualization
- Dealing with model building, data acquisition, cleanup and filtering; identification of patterns & feature extraction, variable selection, experimentation with algorithms and performance analysis

### Software Engineer

**Infonam**

📅 04/2015 - 10/2015    📍 Ho Chi Minh

### Software Engineer

**ISePRO**

📅 02/2014 - 04/2015    📍 Ho Chi Minh

## EDUCATION

### Master of Computer Science

**John von Neumann Institute - Vietnam National University**

📅 09/2015 - Ongoing    📍 Ho Chi Minh

### Bachelor of Information Technology

**University of Science - Vietnam National University**

📅 09/2010 - 07/2014    📍 Ho Chi Minh

## QUOTES

*In order to succeed, your desire for success should be greater than your fear of failure*

Bill Cosby

## SKILLS

**Machine Learning**

**Feature Selection**

**Python**

**Text Mining**

**Ensemble Learning**

**Research and Development**

**Leadership**

**Computer Network**

## LANGUAGES

**Vietnamese**
Native

**English**
Advanced

# PROJECTS

## Automatic labeling for topic models

Develop company product's features

- Find state-of-the-art method for labeling the topics in topic models by performing literature review and experimenting with real data
- Work closely with Data Engineer and Full Stack Development team to deploy the machine learning model to production

## Automated Trajectory Clustering

Master thesis

- Cluster the flight traffics into typical route, the clustering result can be used for flow optimization and measure performance
- The techniques involved: desity based clustering, data interpolating and sampling, curve distance measuring, and visualization

## Customer Churn Analysis

Customer Project

- Employ visual analysis to extract the insight from human interaction
- Hepl client to do identify the factors that influence to customer churn and build predictive model to predict the churn

## Network Graph Analysis

Customer Project

- Detect the influencer in the network graph
- Predict the employee churn based on their interaction in the network

## Sentiment Classification

Customer Project

- Improve accuracy for sentiment classification model and collborate with stakeholders to bring machine leaning model to production
- Using technology: pandas, sklearn, gensim, nltk, doc2vec, matplotlib
- Transform text into numberic vector by using word embedding , and leverage multiple algorithms and dectect the best method suits client's data characteristic

## Text Classification

Customer Project

- Help the client to improve the machine learning model performance in term of accuracy and time
- Dealing with multi-class classification (15 classes), imbalanced data

## San Francisco Crime Classification

Kaggle Competition

- Practice the data science skills by applying the KDD process for data analysis, achieved top 17% in the leaderboard

# TECHNOLOGIES

## Data Science

sklearn    xgboost    pandas

spark    matplotlib    tableau

nltk    git    h2o    mysql

mongodb    numpy    scipy

pattern    aws    jira    intellij

gensim    networkx    linux

## Network Security

nmap    vmware    kerberos

truecrypt    acunetix    openssh

# REFERENCES

**Dr. An Mai**
an.mai@jvn.edu.vn

# CERTIFICATES

**Statistical Learning**
Stanford Online

**Structuring Machine Learning Projects**
Coursera

**AWS Essential**
Linux Academy

**IDS/IPS**
Cybrary