



# Nguyen Thi Thuy Duong

Data Scientist

**Phone:** 0945704548  
**Email:** thuyduong181298@gmail.com  
**Address:** Ho Chi Minh, Vietnam  
**Website:** <https://github.com/Duongntt181298>

## EDUCATION

Sep 2016 - Oct 2020	<b>Faculty of Mathematics and Computer Science, Ho Chi Minh University of Science</b> Major: Probability and Statistics GPA: 3.19/4
---------------------	---

## WORK EXPERIENCE

April 2021 - Present	<b>FPT TELECOM</b> Data Scientist <ul style="list-style-type: none"><li>- Research and offer solutions to help optimize network operations with Machine Learning.</li><li>- Preprocessing big data and build data pipeline to build data lake, data centralize.</li><li>- Analyze and visualize data with Power BI.</li></ul>
May 2020 - April 2021	<b>ESTUARY SOLUTIONS</b> Data Analyst <ul style="list-style-type: none"><li>- Define key metrics and their relationship to measure business success.</li><li>- Review data, design reports and dashboards to meet clients' needs.</li><li>- Execute data modeling and data visualization.</li><li>- Help our business partners understand metric trade-offs and drive influential decisions.</li></ul>
Jan 2020 - Mar 2020	<b>QT DATA GROUP</b> Data Analyst Intern <ul style="list-style-type: none"><li>- Utilize statistical analysis in Python to checking, collecting, cleaning data.</li><li>- Visualize data by using professional Excel chart and supported libraries in Python (Matplotlib, Seaborn).</li></ul>

## SKILLS

<b>Analytical Tools</b>	Spark, R, Excel
<b>Visualization Tools</b>	Power BI, Tableau, Google Data Studio
<b>Programming</b>	Python, SQL
<b>Communication</b>	Communicate my findings with team members and operations team who don't have the same professional background as me.

## PROJECTS

1. DIDIDI PROGRAM
- Data Modeling

- Build model to figure out which is the most successful promotional campaign to bring more benefits by using various algorithms supported by Scikit-learn (Clustering, Regression, Classification, PCA,...).

- **Scorecard**

- Design daily, weekly, monthly and quarterly report in order to monitor clients' business health.
- Implement pipeline with airflow to automatically export report files and notify stakeholders via Telegram and email.

- **Fraud detection**

- Combine multiple algorithms of Scikit-learn (PCA, Isolation Forest) to construct a model that helps detect cheat action.
- Design dashboard to visualize detection result.

- **Report**

- Calculate basic points, bonus points, year points of business partner's.
  - Verify report files by using SQL and Python to ensure accuracy.
  - Implement pipeline to automatically export redemption, reset ranking, reserve points files.
- 

## 2. DOWNTIME

- **Data Modeling**

- Build a system to collect data on customers who lose connection in +- 5 minutes and collect data on customers who reconnect after closed/resolved ticket in +- 120 minutes.
- Analyze data, set up a formula to calculate customer downtime based on the time the customer lost connection.

- **Dashboard**

- Visualize customer downtime by customer group, ticket issue to help operations team helps the operation team to monitor and support technical team in troubleshooting.
- 

## 3. INCIDENT FORECAST

- **Data Modeling**

- Collect temperature data, build a model to predict the temperature of network devices within the next 6 hours with LSTM model.
- Classifies devices at risk of failure by the temperature just predicted with the analyzed temperature threshold.

- **Monitor**

- Build dashboards to monitor real-time events predicted and actual temperatures.
  - Shoot a warning ticket on the incident monitoring system to check and handle it before the device has a problem.
- 

## 4. MODULES

- Fraud detection applied in VoIP system.
  - Spam message detection applied in OnCX system.
  - Eye tracking applied in online learning monitoring system.
-