

# LONG (Lucas) DUC NGO

Canterbury, Sydney, Australia 2193

LinkedIn: <https://www.linkedin.com/in/ngoduclong46>

T: (+61)466043863 Email: [ngoduclong46@gmail.com](mailto:ngoduclong46@gmail.com)

GitHub: <https://github.com/lucasngo>

## PERSONAL PROFILE:

Experienced Data Scientist skilled in machine learning, predictive modeling techniques. Proficient in Python, R, SQL, and cloud technologies (AWS, Azure). Adept at designing scalable data models, optimizing ML pipelines, and transforming complex data into compelling narratives that drive strategic decisions. Passionate about leveraging AI to uncover insights, solve real-world problems, and enhance business outcomes.

## PROFESSIONAL SKILLS and ATTRIBUTES:

- Proficiency in Python, R and experience working with databases and complex queries (PostgreSQL, MongoDB).
- Experience working with cloud services (AWS) and complex technical stack.
- Competence in developing and improving Machine Learning, Deep Learning models.
- Proficiency in data visualization in visualization tools (Tableau, Power BI) and Python packages

## WORK EXPERIENCE:

### **Oct 2022 – Present**

#### **Data Scientist (Full-time)**

**Dataro, Sydney**

##### Key Responsibilities:

- Develop, maintain and improve ML models, pipelines and systems.
- Analyse data systems, ML models and produce reports to support business objectives and CS queries.
- Assist clients through the onboarding process including QA of ML models & predictions and related tools.
- Design and develop dashboards and reports to monitor clients' models and provide actionable insights for proactive intervention.

##### Contributions:

- Generalise and improve 'tagging' model accuracy to 95% while reducing the number of required data points using active learning ML.
- Develop a Machine Learning model to detect potential Regular Givers increasing clients' ROI by 40%
- Develop an age and gender prediction model based on name, donation and communication activities.
- Leverage Generative AI model to automate data transformation, reducing client processing time by 80%.
- Construct dashboards to help detect clients' data/model issues.

Technologies used: Python, SQL, AWS (S3, Athena, Sage Maker, Lambda, CloudWatch, etc), Machine learning, Generative AI.

### **May 2021– Oct 2022**

#### **Junior Data Scientist (Full-time)**

**Skwill, Sydney**

##### Key Responsibilities:

- Help business stakeholders interpret data to enable them to derive meaningful, valuable insights.
- Design and improve new or existing statistical or mathematical software programs in support of research projects.
- Automate data transformation, storing, analyzing, and creating reports about the team's performance based on predefined standards.
- Develop a model to analyze Job Description mapping with a predefined Individual Performance matrix.

##### Contributions:

- Generate valuable insights and recommendations for performance diagnostics with a high acceptance rate.
- Analyze personal performance using statistical and mathematical models.

- High code acceptance rate with less error, less rework and easy to understand.
- Develop the most viable product (MVP) for a successful seeds round.

Technologies used: Python, JavaScript, Machine Learning, NLP, MongoDB, Azure

#### August – November 2020

##### DATA SCIENCE, INTERN (Part-time)

AI Australia, Sydney

##### Key Responsibilities:

- Develop the algorithm to filter potential candidates which is used for the applications screening process.
- Collect data from different open sources and apply data manipulation and engineering techniques to transform raw data into desired formats.
- Perform analysis, building portfolio and modelling the spread of Covid-19.

Technologies used: Python, R, Power BI, Power App, Selenium, Machine Learning

#### June – August 2019

##### DATA SCIENCE, INTERN (Part-time)

Vietnam Prosperity Joint-Stock Commercial Bank (Hanoi)

##### Key Responsibilities:

- Construct a Machine Learning model to detect potential fraudulent transactions.
- Operate data analysis to find patterns and create recommendations system to suggest potential customers and predict system failure.

Technologies used: Python, MySQL, AWS, Machine Learning

## EDUCATION

#### 2018 – 2021 Bachelor of Data Science - S P Jain School of Global Management, Sydney

(Distinction – GPA: 3.4)

Coursework: Introduction to Data Science, Databases, Statistics, Machine Learning, Data Mining, Data Warehousing and Integration, Data Structure and Algorithm, Simulation Modelling, Data Analytics, Visual Analytics, NoSQL, Calculus, Linear Algebra, Big Data.

#### 2015 – 2018 Nguyen Hue High School for gifted students, Hanoi, Vietnam (specialization in math)

(GPA:8.9/10)

## CERTIFICATIONS

- AWS Cloud Practitioner Certificate (valid till Jan 2027)

## APPLIED ACADEMIC and PERSONAL PROJECTS

- **Beijing housing price prediction platform** (Machine Learning, Django, JavaScript, ORM, PostgreSQL).

Build a web application to predict house prices, and find similar houses based on given characteristics with personalized search, and prediction history.

Repository: <https://github.com/lucasngo/Beijing-housing-price>

- **Smart Traffic light system using Computer Vision with web-based control panel and simulations.**

(Computer Vision, Django, Channels).

A vehicle traffic control system that uses computer vision in the backend to detect traffic load in each lane and adopt algorithms to intelligently route the traffic and the time allotted for each green light. Simulation is used to test and find the best algorithm while the web-based control panel is for displaying and interaction.

Repository: <https://github.com/lucasngo/Smart-Traffic-Signal-Using-Computer-Vision-with-control-panel>

- **Machine Learning algorithm comparison on financial data app** (Machine Learning, Dash, API)

Create a web application for users to select stocks and apply various Machine Learning algorithms on these financial data. The app's key features are the ability to automatically apply various Machine Learning algorithms with different hyperparameters and return matrices containing the performance of all listed algorithms.

Repository: [https://github.com/lucasngo/ML\\_algo\\_compare](https://github.com/lucasngo/ML_algo_compare)