

Danh is a Microsoft Certified Data Engineer, Data Scientist, and an open-source contributor. Applying Data engineering and Machine Learning to derive actionable insights is his passion. Danh has experience in collaborating with key stakeholders and acquiring domain knowledge to frame and solve business problems. He sharpens his skills in full stack data science and engineering cycles, which include building data pipelines (ETL), as well as developing and productionizing Statistical and Machine Learning models.

## EXPERIENCE

### Data Scientist - Data Analyst, Johnson Controls, Australia

Oct 2021 - Present

Johnson Controls is a S&P500 firm providing smart building solutions. Danh works across different teams and communicates effectively with various stakeholders. He helps the firm to define business pain points and then forms innovative data science and engineering solutions.

- Build data pipelines (ETL), perform data analysis, data modelling, and develop high quality Business Intelligence (BI) reports using SQL, Python and Power BI.
- Address data inconsistency issues by mapping different data sources and improve data quality.
- Develop machine learning models and A/B experiments; deploy models on Microsoft Azure.

### Open-source Software Contributor, PyMC on GitHub

Jan 2022 - Present

- Add a new [Multi-output Gaussian Processes](#) feature for time series modelling into PyMC
- Accelerate the [CI pipelines](#) for PyTensor to reduce testing time by **60%**, from 75 to 26 minutes.
- Develop features, fix issues, CI/CD and docker images for open-source projects that focus on probabilistic Machine Learning, including [PyMC](#), [PyMC-Examples](#), and [PyTensor](#).

### Data Associate Instructor, Data Fluency at Monash University, Australia

Jul 2021 – Dec 2022

- Prepare technical materials; Teach data-related [hands-on workshops](#) on Python, Git, Bash, High Performance Computing, and Power BI for researchers at Monash University.

### Data Analyst Intern, Domain Group, Australia

Aug 2018 - Nov 2018

- Developed [a chatbot](#) and customer clustering models, and built dashboards with Tableau.
- Performed data analysis using SQL and Python on AWS RedShift and Snowflake.

## CERTIFICATES

- [Microsoft Certified: Azure Data Engineer Associate](#) - Microsoft Dec 2022
- [Microsoft Certified: Azure Data Scientist Associate](#) - Microsoft Sep 2022
- [AWS Certified Cloud Practitioner](#) - Amazon Web Services Oct 2022
- [Deep Learning Specialization: five courses](#) - Coursera Sep 2021
- [Discrete Choice Analysis: Predicting Market Demand](#) - MIT Jun 2021
- [Scalable Machine Learning with Apache Spark](#) - Databricks Jun 2021

## SKILLS

- **Technical:** Python, SQL, Scala, R, Bash, Java, C++, Rust, ETL, ITIL, Cloud, Data Warehouse
- **Data analysis:** NumPy, Jax, Pandas, Plotly, Jupyter Notebooks, Tableau, Power BI, Excel
- **Machine learning:** Scikit-learn, XGBoost, PyMC, Statsmodels, PyTorch, TensorFlow
- **MLOps & DataOps:** MLFlow, Docker, Kubernetes, GitHub Actions, Great Expectations
- **Data Engineering:** SQL, T-SQL, Hadoop, PySpark, Dbt, Kafka, Airflow, Azure Data Factory
- **Environments:** Azure Logic apps, Gen2, Azure SQL, Azure Synapse, AWS, RedShift, Snowflake

## RECENT PROJECTS

For further projects, please visit <https://danhphan.net/projects>

- **Interest rates predictions:** <https://danhphan.net/projects/#interest-rates-predictions>  
This project aims to predict the base interest rates in Australia and the US. I demonstrate a complete cycle in Data Engineering and Machine Learning, including an ETL process to collect and transform data from different data sources.
- **Modelling the performance of baseball players:** <https://danhphan.net/multi-output-gps>  
We model the performances of different sport players by leveraging Multi-output Gaussian processes, which can simultaneously infer many outputs with the same source of uncertainty. This work is supported by [Google Summer of Code](#) 2022.
- **Aussie Social Sentiment Analysis:** <https://danhphan.net/aussie-sentiments>  
This project collects data from Twitter's APIs, then cleans and stores in a SQL database. The data is then used for model training and prediction of sentiment analysis for other tweets for Australians.
- **Visualise Melbournian Daily Activities:** <https://danhphan.net/melbournian-daily-activities>  
This project visualises the daily activities of Melbournians in different areas.

## EDUCATION

**PhD researcher**, Monash University, Australia Feb 2020 – Feb 2023

My research focuses on applying Machine learning to understand what activity people do in the day, where these activities happen, and how people travel among locations.

- Research topic: Machine learning for intelligent transport systems and activity-based models.
- Received Monash Graduate Full Scholarship and Travel Grants.

**Master of Information Technology**, Macquarie University, Australia Feb 2017 - Dec 2018

- Core units: Machine Learning, Data Mining. GPA 4.0/4.0.
- Received Macquarie University Award for Academic Excellence.

**Bachelor of Information Technology**, Hanoi University of Technology, Vietnam 2005 - 2010

- Core units: Probability and Statistics, Algebra and Analytic, Numerical methods, Discrete Mathematics, Artificial Intelligence, Data structures and algorithms, Software Engineering.
- High Distinction, GPA 8.3/10, top 3% of the IT Faculty.

## RECENT PUBLICATIONS

For further publications, please visit my [Google Scholar profile](#)

- Danh Phan, Hai Vu, Graham Currie. AttentionChoice: Discrete choice modelling supported by a deep learning attention mechanism. *The 102<sup>nd</sup> Transportation Research Board*, 2023.
- Danh Phan, Hai Vu, Eric Miller. A New Approach to Improve Destination Choice by Ranking Personal Preferences. *Transportation Research Part C: Emerging Technologies*, 2022.
- Danh Phan. Australia Housing Price Prediction using Machine Learning Algorithms: The Case of Melbourne City, Australia. *The 2018 IEEE International Conf. on Machine Learning*, 2018.

## VOLUNTARY WORK

- Assisted PhD researchers on data analysis issues in Data Fluency's Slack Jul 2021 - Present
- Supervised students on final year projects at Monash University Feb 2021 - Dec 2022
- Mentored people working on open-source projects with the PyMC team Jul 2022 - Aug 2022

## REFERENCES

Available upon requests.