# DANG DUY DINH

dddinh@bu.edu | 919 – 746 – 2989 | Portfolio | LinkedIn | Boston, MA

## SUMMARY

Aspiring Data Scientist and results-driven graduate student at Boston University with a unique interdisciplinary background in economics, information technology, and advanced analytics.

## EDUCATION

**Boston University** | Boston, MA — *Expected Graduation: December 2026*
Master of Science in Data Science

**North Carolina State University** | Raleigh, NC — *Graduated: May 2025*
Bachelor of Science in Business Administration, Information Technology — GPA: 3.93
Bachelor of Arts in Economics
**Relevant Coursework:** Data Analytics, Database Management, Deep Learning, Tools for Data Science

## PROJECTS

**Campaign Zero – 911 Call Data Patterns**
*(Tech Stack: Pandas, Snowflake, Data Pipeline)*

- Collaborating with Campaign Zero, a national nonprofit dedicated to ending police violence through data-driven policy reform, to build a scalable pipeline for 911 call data aggregation across 11 major U.S. cities.
- Utilizing Python for data wrangling and pipeline development, Snowflake and Azure for cloud-based ETL automation, and Tableau for designing interactive, policy-relevant dashboards.
- Delivering a harmonized, analysis-ready dataset and an actionable final report to support policy recommendations and future integration into Campaign Zero's Safe Cities platform.

**Tree-or-Not Classifier with Deep Learning Model**
*(Tech Stack: PyTorch, Neural Network, Regularization)*

- Diagnosed performance bottlenecks in baseline CNN architecture for Tree-or-Not dataset (70% validation accuracy, 68% test accuracy) attributed to initialization and regularization deficiencies.
- Designed and executed systematic hyperparameter search across 16 configurations testing different initialization schemes and regularization techniques to address overfitting and improve generalization.
- Achieved 12-point validation accuracy improvement (70%→82%) and 7-point test gain (68%→75%); diagnosed remaining validation-test gap and documented strategies to further improve model generalization.

**Stroke Prediction – Kaggle Challenge**
*(Tech Stack: Python, Scikit-learn, Pandas)*

- Performed detailed exploratory data analysis (EDA) on the StrokeX dataset to identify trends, detect outliers, and uncover correlations between demographic and health-related features.
- Developed, tuned, and evaluated multiple classification models (including logistic regression and ensemble methods) to predict stroke occurrence, achieving a final test accuracy of 96%.

## EXPERIENCE

**Network Operation Center Intern | OIT ComTech** — *May 2024 – May 2025*

- Resolved 20+ weekly technical tickets involving VoIP system issues and network device configurations.
- Developed and documented a standardized workflow for updating records across Call Manager, Unity Connections, and E911 systems, improving process speed and reducing manual errors.
- Partnered with network analysts, configuring over 60 Cisco switches, guaranteeing optimal configuration of VoIP and data devices, resolving network bottlenecks and improving data transmission speeds.

**RAG AI Researcher | NC State University** — *October 2024 – May 2025*

- Researched on RAG model and conducted performance evaluations across LLM platforms.
- Designed experiments with metrics to assess performance, latency, and accuracy of RAG-powered prototype chatbots for educational use cases
- Delivered a report each biweekly meeting with system recommendations, contributing to enhancements adopted by technician team that reduced runtime by 19%.

## SKILLS

- **Programming Language**: Python, R, SQL
- **Tools**: Snowflake, Tableau / Power BI, Stata, Git, Jupyter Notebook, Excel
- **Techniques**: Machine Learning, Data Mining, Statistical Analysis, Data Visualization, Feature Engineering

## CERTIFICATION

- Python for Data Science (*Codecademy*)
- Intro to Snowflake for Data Scientist (*Coursera*)