# Ngu Dang

Boston, MA, 02215 — dsngu97@gmail.com — nathandang97.github.io — linkedin.com/in/ngu-dang — github.com/NathanDang97

## Education

**Boston University** **09/2020 - 05/2026 (expected)**
*Doctorate in Computer Science*
- **GPA:** 3.93/4.00 – Passed the PhD Candidate Qualifying Exam.
- **Relevant Coursework:** Machine Learning, Natural Language Processing, Neural Networks and Deep Learning (DeepLearning.AI Professional Certificate), Data Science (IBM Professional Certificate).

**Clark University** **01/2018 - 05/2020**
*Bachelor of Arts in Computer Science*
- **Minors:** Data Science, Mathematics
- **GPA:** 3.93/4.00 – First Honors Dean's List in 2018, 2019, and 2020.
- Graduated with *Summa Cum Laude* and obtained *Outstanding Academic Achievement Award in CS*.

## Skills

**Programming:** Python, Java, C, C++, MySQL.
**Libraries:** Pandas, Numpy, PyTorch, TensorFlow, Keras, Scikit-learn, Seaborn, NLTK.
**Script:** HTML, CSS, LaTeX.
**OS/Tools:** Linux, Windows, Git, Jupyter, Google Colab, Visual Studio, Microsoft Office Suite.

## Projects

**Churn Predictor for Subscription Services** **03/2025 – 04/2025**
- Implemented an end-to-end churn prediction pipeline in Python for a video streaming service using a real-world imbalanced subscription dataset using an ensemble of three models — a neural network, XGBoost, and Random Forest — using weighted soft voting to optimize class ranking and maximize AUC.
- Engineered advanced features (e.g., ratio metrics, interaction terms, log transforms, and behavioral buckets) on top of 20 given features to boost signal quality and improve model discrimination.
- The model achieved a ROC AUC score of 0.75 (91st percentile). Github link.

**Edible Mushroom Classifier** **07/2024 – 08/2024**
- Implemented a Random Forest model classifying edible mushrooms from toxic ones in Python based on their physical properties and characteristics such as the shape, size, and color of the mushroom's cap and stem, etc. which yields a good mix of numerical and categorical features for the data.
- The dataset used in this project (train and test) was generated from a deep learning model trained on the UCI Mushroom dataset. The training set and test set contained 3116945 and 2077964 data points respectively with 22 features.
- The model achieved an accuracy score of 0.987. Github link.

**Digit Recognizer** **10/2023 – 11/2023**
- Implemented a Digit Recognizer model in Python using Convolutional Neural Network (CNN) trained on the MNIST dataset consisting of handwritten digits (0-9).
- The model achieved an accuracy score of 0.988. Github link.

**House Price Predictor** **07/2022 – 08/2022**
- Implemented a House Price predictor model using CatBoost Regression where the data contained 2919 entries, each with 79 explanatory features describing most aspects of residential homes in Ames, Iowa, such as number of bathrooms and bedrooms, utilities, locations, front-yard and backyard condition, etc. which yields a good mix of numerical and categorical features for the data.
- The model achieved an RMSE score of 0.13. Github link.

## Experience

**Boston University** **09/2020 – present**
*Graduate Research & Teaching Assistant*
- Working on joint projects with professors, postdocs, and PhD students where we study the computational lower bounds and upper bounds of complex algorithms and design improvements on top of current state-of-the-art results.
- Implementing our algorithms, experiments, and demo scripts in Python.
- Leading weekly discussion/lab sections, office hours, for Algorithms, Theory of Computation, and fundamental math classes such as Discrete Math, Linear Algebra, and Probability.

**Clark University** **05/2019 – 05/2020**
*Undergraduate Research Assistant*
- Contributed to computer vision and computational geometry research projects for the Computer Science Department.
- Implemented experiments, statistical analysis, visualization, and geometrical simulations in Python and Java.