

# Ninh Giang Nguyen - Data Scientist/Analyst

4000 MayFlower Hill, Waterville, Maine | 207-660-7715 | [ngnguy26@colby.edu](mailto:ngnguy26@colby.edu) | [LinkedIn](#) | [Portfolio](#)

## EDUCATION

**Colby College**, Waterville, ME, USA, *Bachelor's Degree*

**Graduating May 2026**

**Major:** Computer Science & Economics **Minor:** Statistics

**Major GPA:** 3.91/4.0

**Relevant Coursework:** Data Analysis & Visualization, Statistical Modeling, Statistics in Data Science, Financial Technology, Probability, Neural Networks, Computer Vision, Data Structure & Algorithms, Software Engineering, Declarative programming,

## TECHNICAL SKILLS & INTERESTS

**Technical Skills:** Python (Pandas, NumPy, Scikit-learn, TensorFlow, SHAP, Joblib), SQL (Hive, Spark), R, Java, AWS, APIs, Snowflake, Git

**Data & Statistical Tools:** Predictive Analytics, Big Data Query, Excel VBA, Seaborn, Tableau, PowerBI, Data Mining, Statistical Inference

**Machine Learning & AI:** PyTorch, Keras, Clustering, PCA, Model Evaluation, Neural Networks, Deep Learning, Reinforcement Learning

## WORK EXPERIENCE

**VPBank**, Hanoi, Vietnam

**June - September 2025**

Data Science intern

- Analyzed and processed 5M+ customer profiles from VPBank's app and website to develop a Credit Card Approval prediction model, identifying profiles with the highest probability of successful registration and fund disbursement
- Conducted exploratory data analysis (EDA) and feature engineering using **SQL** and **Big Data Query** tools, developed and trained models leveraging advanced AI techniques (DeepGBM, SAINT, IGTD), rigorously benchmarked model performance, and deployed the optimal solution on **AWS EC2**.
- Designed presentation and delivered project defense using **PowerBI** to 10+ executive business analysts and stakeholders
- Researched advanced applications such as AI-driven risk scoring and fraud detection, collaborating with senior data scientists

**Greentech Vina**, Hanoi, Vietnam

**June - September 2024**

Data Analyst

- Consolidated and refined 10M+ international trade records between Vietnam and 90+ other countries on Ziploc and Zipper bag products, leveraging **Excel** and **Python** to optimize product distribution and forecast market
- Performed trend and pattern analysis with **Tableau** that identified anomalies, price shifts and key customer segments for data-driven decision-making, contributing to a projected 15% increase in engagement rates
- Streamlined cross-department collaboration between analytics, sales, and marketing teams by redesigning reporting pipelines, cutting manual analysis, and improving insight turnaround time by 30%

**Colby College**, Waterville, ME, USA

**September 2023 - current**

Head of Math Teaching Assistant

- Facilitated study sessions for 30+ students, effectively clarifying complex concepts in Linear Algebra and Statistical Inference
- Coordinated with faculty to grade and provide detailed, constructive feedback, ensuring accurate academic evaluation

## PROJECTS

**FutureProof: Bankruptcy Detection** | [Github](#)

**May 2024**

- Built predictive models to estimate bankruptcy risk for ~7,000 companies using financial performance metrics
- Applied **Joblib**, **Seaborn**, **scikit-learn**, and **SHAP** to conduct EDA, feature engineering, data and results visualization, model behavior interpretation and feature importance
- Optimized tree-based models (**XGBoost**, **Random Forest**, **SVM**, and **LightGBM**, **CatBoost**), achieving >90% accuracy
- Deployed an interactive dashboard leveraging **Streamlit** to predict bankruptcy and highlight top 5 financial risk indicators

**PharmaGuard** | [Github](#)

**January 2025**

- Engineered a **MySQL database** with three interlinked schemas (**APIs** from OpenFDA, DailyMed, MedlinePlus) to integrate drug details, recalls, and adverse events into a comprehensive system, enabling sub-second query performance
- Compiled a **Python-based ETL pipeline** with **Pandas**, **SQL** and **Spark** to extract, clean, and store drug information, and utilized **Hive** for efficient querying and up-to-date aggregation of large datasets

**AdoptGraph: Social Tech Uptake Forecasting** | [Github](#)

**June 2025**

- Forecasted technology adoption among Facebook users by analyzing 100K+ social graph nodes and edges, incorporating features such as friend count, interaction frequency, and neighbors' behavior
- Implemented a 3-layer **Graph Convolutional Network** with **PyTorch Geometric**, reaching a baseline accuracy of ~80%