# Vuthtyra (Teera) Yong

Ithaca NY 14850 | (607) 229-889 | vuthtyrayong@gmail.com | https://ty3340.github.io | https://www.linkedin.com/in/vuyong/

## EDUCATION

**Syracuse University,** School of Information Studies
M.S. in Applied Data Science, **GPA: 4.0/4.0** **December 2025**

**Cornell University,** College of Human Ecology, Ithaca, NY **August 2021**
M.S. in Fiber Science, **GPA: 3.97/4.3**

**University of California, Davis,** College of Engineering, Davis, CA **August 2018**
B.S.E. in Materials Science and Engineering**, GPA: 3.44/4.0**

*Relevant Coursework:* Applied Machine Learning • Natural Language Processing • Cloud Management • Advanced Big Data Management • Big Data Analytics • Data Warehousing • Intro to Data Science • Business Analytics • Quantitative Reasoning Data Science • Data Admin & Database Management

## SKILLS

Programming: Python, R, SQL, C++
Machine Learning Techniques: Supervised & Unsupervised Learning, Deep Learning, Time Series Forecasting
Data Visualization & BI Tools: Tableau, Power BI, Excel, Access, Google Analytics, Power Query, Power Pivot
Cloud Platforms: Azure Cloud, AWS Cloud, Snowflake
Big Data & Databases: MongoDB, Redis, Cassandra, MiniO, Hadoop
Data Engineering Tools: dbt Labs, Apache Spark, Apache Hadoop
Storage Systems: Amazon S3, Minio, HDFS/Hive
DevOps & Other Tools: Docker, Kubernetes, Git & Version Control, Linux

**Language**: Khmer (native), English (professionally fluent)

## RELEVANT ACADEMIC PROJECTS

Amazon Product Reviews 2023:
- Transformed 153,000 Amazon product reviews into 768-dimensional vectors using BERT embeddings
- Applied dimensionality reduction techniques (UMAP, t-SNE, PCA) to visualize high-dimensional embeddings and uncover hidden patterns in review sentiment and product categories.
- Performed unsupervised clustering using K-Means to identify the optimal number of clusters in BERT-based embeddings
- Leveraged GPT-4 to automatically generate descriptive labels for 90 clusters, and summarized the positive and negative aspects within each cluster to extract actionable insights.

Airbnb Fraud Detection:
- Developed a fraud detection model using a Feed-Forward Neural Network (FNN) trained on the Boston Airbnb dataset
- Scaled and deployed the model using Docker and Kubernetes, and hosted it on Azure Cloud.

Household Energy Forecasting:
- Conducted feature engineering and exploratory data analysis to identify key drivers of peak energy demand, such as temperature, house size, and time-of-day.
- Built and tuned an XGBoost regression model to predict hourly electricity usage during July, achieving 72% accuracy using static housing, weather, and time-series energy data from 5,000 South Carolina homes (2018).

## CERTIFICATIONS & BADGES

AWS Certified Cloud Practitioner **February 2025**
Microsoft Certified: Azure Fundamentals **February 2025**
Snowflake Hands-on Essentials: Data Warehouse **January 2025**