

Thomas Lee

Seattle, WA | (206) 493-3040 | lee.minghsieh@northeastern.edu

[linkedin.com/in/mhleethomas](https://www.linkedin.com/in/mhleethomas) | github.com/mhleethomas | [mhleethomas.github.io](https://github.io/mhleethomas)

SUMMARY

Data-driven professional with expertise in statistical analysis, process optimization, and automation. Skilled in designing experiments, leveraging statistical models, and implementing innovative solutions to improve processes and outcomes. Passionate about leveraging machine learning to drive business insights and efficiency.

SKILLS

- **Programming & Tools:** Python, Java, SQL, Visual Basic, Git, GitHub, Jupyter Notebooks, DBeaver
- **Machine Learning:** Scikit-learn, TensorFlow, Keras, XGBoost, Pandas, NumPy, Statsmodels
- **Databases & Data Engineering:** SQL Server, PostgreSQL, MongoDB Atlas, ETL Pipelines
- **Data Visualization:** Tableau, Matplotlib, Seaborn, Excel, JMP

EDUCATION

Northeastern University - Seattle, WA

Expected Jun. 2026

Master of Information Systems

- Relevant Coursework: Data Science Engineering, Database Design, Program Structure and Algorithms

National Yang Ming Chiao Tung University - Hsinchu, Taiwan

Sept. 2017 - Jun. 2021

Bachelor of Applied Chemistry

- Relevant Coursework: Analytical Chemistry, Integrated Chemistry Laboratory, Calculus

WORK EXPERIENCE

Kinsus Interconnect Technology Corp. - Hsinchu, Taiwan

May. 2022 - Sept. 2023

Process Engineer, Process Technology Engineering Department

- Conducted statistical analysis and Design of Experiments (DOE) to optimize production processes, achieving a 100% reduction in defect rate and enabling mass production, which earned the Star Award for excellence
- Developed and implemented automated measurement systems integrated with Statistical Process Control (SPC) databases, improving measurement accuracy by 20-fold and capacity by 14-fold
- Leveraged ETL techniques (data extraction, transformation, and validation) to preprocess accuracy metrics such as Gauge Repeatability and Reproducibility (GRR), significantly reducing measurement errors
- Applied Excel Macros and VBA scripting to optimize process parameters, enhancing sorting accuracy and improving defect detection capabilities in production workflows

PROJECTS

Surface Crack Detection in Concrete Using Deep Learning

- Developed a Convolutional Neural Network (CNN) to classify concrete surfaces as cracked or uncracked, utilizing preprocessing techniques such as resizing, augmentation, and normalization on a dataset of 40K+ images, achieving 99.89% accuracy
- Optimized model performance with hyperparameter tuning, resulting in a model with 1M+ parameters
- **Technologies:** Python, TensorFlow, Keras, NumPy, Pandas, Matplotlib, OpenCV

House Price Prediction with Statistical and Geographical Data

- Conducted data cleaning, feature engineering, and visualization using GeoPandas and Seaborn to analyze spatial and statistical relationships in housing data
- Built and optimized regression models, achieving an RMSE of 0.15 (log-transformed scale) and R² of 0.91
- **Technologies:** Python, Scikit-learn, XGBoost, NumPy, Pandas, GeoPandas, Matplotlib, Seaborn