

Minhyeok (Shawn) Son

shawn22587@gmail.com | linkedin.com/in/minhyeokson | github.com/Shawn-Son

EDUCATION

University of Michigan, Ann Arbor

Expected May 2027

Master of Science in Data Science

Ann Arbor, MI

- **Coursework:** Data Structures, Algorithms, Databases, Machine Learning

Seoul National Univ. of Science & Technology

Feb 2025

B.S. in Industrial and Information Systems Engineering

Seoul, South Korea

EXPERIENCE

eBay

May 2026 – Aug 2026

Machine Learning Engineer Intern

San Jose, CA

- Enabled analysts to track **25** types of compliance alerts from a single query via an LLM workflow in Python.
- Automated root-cause analysis by linking **3** kinds of code changes (configs, refreshes, deploys) to anomalies.
- Cut engineer escalations by **80%** (**10 hrs/week**) by letting analysts debug issues themselves.
- Selected the production setup by benchmarking **15** model/prompt configs on accuracy, latency, and cost.

American Airlines

Sep 2025 – Nov 2025

Student Machine Learning Engineer

Ann Arbor, MI

- Unified **1M+** flights into a PostgreSQL feature store by joining operations, weather, and congestion data.
- Captured key delay drivers by engineering **50** weather, schedule, and airport-traffic features.
- Beat baselines by **8 AUC pts** on held-out flights by training and tuning gradient-boosted tree models.
- Refined features and evaluation over **5** review cycles with American Airlines operations staff.

Industrial AI Lab, Seoul National Univ. of Science & Technology

Sep 2022 – Dec 2024

Research Assistant

Seoul, South Korea

- Detected rare defects in a **2.9M**-row sensor stream from a heat-treatment line using XGBoost and SHAP.
- Cut inspection cost on **~263K** semiconductor units with a **<2%** defect rate using knowledge distillation.
- Matched high-cost model accuracy with low-cost features; uncertainty sampling beat random by **3 AUC pts**.
- Validated the framework across **30+** repeated trials, contributing methodology and analysis to the published study.

PROJECTS

eBay ML Competition | 3rd of 80 · Solo

eval.ai/web/challenges

- Extracted product attributes from **10K** German listing titles using XLM-R / mDeBERTa with a word-level CRF.
- Finished within **0.002** of first place on the official recall-weighted F3 metric.
- Stabilized noisy-text predictions with **100+** hand-built rules and a second-stage confidence filter.
- Surpassed **eBay's own baseline** by replicating the official scoring function for local validation.

User-Friendly AutoML Platform | Python, Streamlit, SHAP

github.com/Shawn-Son/Auto-ML-Service

- Enabled no-code ML by building a Streamlit app that auto-detects **4** problem types from uploaded data.
- Took users from raw data to trained models in **15 min** by automating data exploration and preprocessing.
- Selected the best model per task by running automated search across **10** candidate algorithms.
- Made results interpretable via **1-click** downloadable reports with SHAP-based feature explanations per model.

PUBLICATION

Heo, J., Son, M., & Shim, J. "Active Inspection with Knowledge Distillation for Cost-Effective Fault Prediction in Manufacturing Process." *Scientific Reports*, 16, 8613 (2026).

TECHNICAL SKILLS

Languages: Python, Java, SQL (PostgreSQL)

ML / DL: PyTorch, scikit-learn, TensorFlow, XGBoost, Transformers / NLP, CRF, time series, knowledge distillation

Data & Tools: pandas, NumPy, SHAP, Streamlit, Git, Docker, Artifactory, Jira

Work Authorization: F-1 STEM — 36 months OPT eligibility; no immediate sponsorship needed