Seonmin Kim

Los Angeles, CA | kimsunmin0227@gmail.com | linkedin.com/in/sunnykimm12 | github.com/nimnusmik

EDUCATION

University of Southern California, Viterbi School of Engineering, Los Angeles, CA Expected May 2028
M.S. in Applied Data Science
Coursework: Machine Learning for Data Science (DSCI 552), Data Visualization (DSCI 554)

Hankuk University of Foreign Studies, Yongin, South Korea Aug 2024
B.B.A., Global Business and Technology; Double Major, Artificial Intelligence Convergence
Valedictorian (1 of 200+) | GPA: 4.22 / 4.5 | Minister of SMEs and Startups Award for academic excellence (2023)

PROFESSIONAL EXPERIENCE

Global Vision Co., Ltd., Seoul, South Korea Jan 2026 - Jul 2026
Data Engineer, Data Platform and Business Intelligence | Live demo (public data): onbid-demo.onrender.com

- Built solo a daily 5-stage Python ETL pipeline refreshing 3,000+ public auction listings from 2 government APIs.
- Recovered 17K bid records (57% of history) that the source API deletes, by collecting them from the public website.
- Added a quality gate that halts the daily run on empty API responses; idempotent upserts make reruns safe.
- Saved a real estate team member 3 hours of daily searches with a dashboard ranking listings against 490K+ market prices.
- Cut bulk Excel uploads from ~10 minutes to under 1 with async Celery ingestion into PostgreSQL.

SEMO, Seoul, South Korea Sep 2025 - Dec 2025
Analytics Lead Intern

- Wrote 4 BigQuery SQL reports (retention, payment conversion) for a Looker Studio dashboard 3 executives read weekly.
- Designed the GA4 event tracking spec and verified the developers' implementation.

PROJECTS

StockMind: Market Data Pipeline and Trading-Volume Forecasting Sep 2026 - Present
Solo Developer | github.com/nimnusmik/StockMind

- Automated an hourly feed of 100K+ Yahoo Finance posts on 15 tickers with retries, backfill, and row-count checks.
- Forecast next-hour trading-volume bursts (PR-AUC 0.688 vs 0.460 baseline) after catching leakage; posts added no signal.

Toss ML Next Challenge: Ad Click-Through Rate Prediction (Dacon) Sep 2025 - Oct 2025
Team Lead (3) | github.com/nimnusmik/toss-ad-click-prediction

- Ranked 70th of ~700 teams (top 10%) in an ad click prediction contest on 10M+ rows.
- Built a chunked Polars preprocessing pipeline and a DCN-v3 model for a 1:52 class imbalance, tracked in W&B.

Movie Recommendation MLOps Pipeline (Upstage AI Lab) Jul 2025 - Aug 2025
Team of 4 | github.com/nimnusmik/mlops-project-mlops-5

- Co-built an Airflow pipeline automating model training and inference on 3 AWS EC2 servers with MLflow and FastAPI.
- Owned the data layer: TMDB API ingestion, PostgreSQL schema design, and S3 Parquet backups.

RESEARCH EXPERIENCE

University of Seoul, Big Data and AI Economics Lab, Seoul, South Korea Jan 2024 - Jul 2024
Undergraduate Research Assistant

- Forecast building-level electricity demand in Seoul to 2050 under SSP climate scenarios with XGBoost and CatBoost.
- Built the 1.2M-row dataset from ~1,100 S-DoT climate sensors, building registry, and housing price data.

TECHNICAL SKILLS

Data Engineering: ETL/ELT, batch processing, data quality checks, data validation, schema design, REST APIs, web scraping
Tools: Python, SQL, BigQuery, PostgreSQL, SQLite, Airflow, Celery, Pandas, Docker, Git, GitHub Actions, CI/CD
Analytics and BI: GA4, Looker Studio, dashboards, data visualization, funnel and cohort analysis
Machine Learning and Cloud: XGBoost, CatBoost, PyTorch, MLflow, W&B, FastAPI, AWS (EC2, S3), GCP

ADDITIONAL EXPERIENCE AND AWARDS

Performance Marketing Intern, SearchM (Sep 2021 - Feb 2022): Google Analytics reporting for client accounts
Upstage AI Lab Bootcamp (May - Dec 2025) | 2nd Place, HUFS DAT Capstone (2023) | Excellence Award, KDATA (2022)