Minseok Oh

Cupertino, CA 95014

408-334-5898 | ohmseok0524@gmail.com | www.linkedin.com/in/scottmsoh | Portfolio

- Senior data Scientist with 10+ years at ORACLE and LG, driving business performance through data science solutions
- MS in Information Systems(AI/ML) candidate at Santa Clara University (Expected June 2025)
- Led high-impact projects in churn predictions, recommendations, time-series analysis, financial analysis, demand forecasting
- Expert in A/B testing, experimental design, NLP, web crawl, and cloud-based Data Warehouse / ML solutions

SKILLS

PROGRAMMING |Python • SQL • Smart PLS

DATA ANALYSIS | Pandas • NumPy • Regular Expressions (re) • Statistical Analysis • Hypothesis Testing • A/B Testing • Experimental Design

MACHINE LEARNING | Scikit-learn • Recommendation engines • Natural Language Processing • Time Series Analysis • PyTorch • Deep Learning

DATA ENGINEERING | Data Pipeline • Data Modeling • Data Warehouse • Spark • Web Crawling (BeautifulSoup, Selenium, Requests, HTML Parsing) • ETL • GIT

VISUALIZATION | Matplotlib • Seaborn • Plotly • Business Intelligence (Tableau, PowerBI) • Spotfire **CLOUD** | OCI • AWS • GCP **PROJECT MANAGEMENT**| Waterfall • Agile **Soft Skills**| Organizational leadership • Problem Solving • Decision Making • Strategic Analysis

EXPERIENCE

FOLLOZE | Machine Learning, Classification, Natural Language Processing, Web Crawling

Data Scientist Intern (Practicum) | Jan 2024 – June 2024, San Mateo, CA

Customer Propensity Analysis for Sales Forecasting and Targeting

- Predicted future customers across 3,000+ customer profiles to enhance sales forecasting and improve customer acquisition strategies
- Engineered key financial and demographic features from 130+ variables, achieving optimal performance with XGBoost using SMOTE, and attained a recall score of 0.86

ORACLE | Project Delivery, Machine Learning, Recommendation engines, Time Series Analysis, A/B testing, Data Warehouse, ETL, Cloud Computing

Principal (L5) Consultant | May 2019 – July 2023, Seoul, Korea

End-to-End Data Pipeline and Purchase-Based recommendations using Collaborative Filtering

- Increased monthly sales by 4.5% through A/B testing of in-store product recommendations with statistical significance (p-value < 0.05, n=42 days)
- Enhanced scalability and efficiency with FAISS indexing and a two-stage reranking process using Matrix Factorization and XGBoost for top-N recommendations

Early Warning for Customer Churn with XGBoost

- Reduced customer churn by implementing automated risk alert system, identifying high-risk customers on a weekly basis
- Developed customer churn prediction model using XGBoost achieving 0.81 recall score through feature engineering, PCA, and imbalanced data handling

Biosimilar Demand Forecasting Using ARIMA Time Series Analysis

- Led EU/US biosimilar sales forecasting initiative, improving revenue planning accuracy for $300M+ product line
- Developed SARIMAX model improving forecast accuracy from 0.42 to 0.84 via seasonal optimization

EDUCATION

Santa Clara University
M.S., Information Systems (ML/DL)
June 2025 | Santa Clara, CA

Korea Advanced Institute of Science Technology (KAIST)
M.S., Information Management
Feb 2019 | Seoul, Korea

Dongguk University
B.S., Electronic Engineering
Feb 2013 | Seoul, Korea

HONORS & AWARDS

DEAN's List | Honoree for academic excellence, Q3 2024 from SCU

COURSEWORK

Machine Learning
Deep Learning
Natural Language Processing
Reinforce Learning
Big Data Analysis
Data Analytics
Business Analytics
Information System Design-Analysis
Database Management Systems
Software Project Management
Database Design
DW&BI
Object-oriented Programming
Mathematics, Statistics

PUBLICATION [PDF]

"Analyzing Expected Values on the Continuous Intention to Use IoT Products Based on Expectation-Confirmation Theory"
- Tailored business strategies are needed based on income, household size, and region
- PLS-SEM and Bootstrap analysis were used to validate relationships and evaluate model significance.

LG DISPLAY | **Natural Language Processing, Deep Learning, Machine Learning, Web Scraping**

Data Scientist | January 2013 – May 2019, Seoul, Korea

News Sentiment Analysis System for IR Strategy

- Enhanced investor relations by analyzing sentiment across 10K+ monthly news articles, driving data-driven disclosure decisions
- Developed a sentiment analysis engine with ElasticNet (0.80 R square), using BeautifulSoup, Requests, and asynchronous scraping to process 400+ daily articles efficiently.

Machine Fault Detection Classification

- A fault detection system that enables prompt maintenance response by monitoring current machine conditions
- A machine fault detection model with feature scaling and imbalance handling, achieving F1-score of 0.87

KAIST DATA RESEARCH LAB | **A/B Testing, Recommendations, Data mining**

Researcher | September 2016 – January 2019, Seoul, Korea

E-commerce Recommendation Engine: Matrix Factorization with Implicit Feedback and Negative Sampling

- Increased Click-Through Rate (CTR) by 20% for e-commerce product recommendations by leveraging user click history
- Developed Matrix Factorization model with implicit feedback and negative sampling techniques, optimizing BPR score 0.78 with FAISS-based efficient retrieval system

Marketing Analytics - Facebook Ads Optimization

- Design and run time zone/age-based A/B tests for female-targeted Facebook ads| 12% improvement in CTR through target segment optimization (Lift 1.12, p-value < 0.05, 30k samples)

Marketing Analytics - Campaign Performance Optimization

- Conducted A/B tests targeting professionals with tailored startup and recommendation-focused content, achieving a 3.4% increase in CVR (p-value < 0.05, 5.2k samples)

SCU PROJECT | **Machine Learning, Natural Language Processing, Web Crawling, Large Language Model, Artificial Intelligence, Recommendation Systems**

Participant | September 2023 – Present, Santa Clara, CA

Financial QA & Sentiment Analysis chatbot

- Developed a financial news analysis chatbot integrating fine-tuned QA (Phi-2) and sentiment analysis (RoBERTa) models, enabling users to extract actionable insights and make informed financial decisions efficiently

Movie recommendation engine with GraphSage algorithms and similarity search

- Engineered movie recommendation engine leveraging graph neural networks (GraphSage) and efficient similarity search (Annoy)

Certificates [PDF]

Oracle Cloud Analytics Professional,
Autonomous Database Cloud Specialist,
Cloud Infra Architect Professional,
Cloud Infra Architect,
Cloud Infra Foundation