

Data Science Portfolio

Selected Works for Senior Data Scientist Position at Geotab

Seung-il Lee

Seoul, South Korea | lsi429@naver.com | <https://drfirstlee.github.io/>

1. IoT Telematics Data Pipeline & Driver Safety Scoring

Data Scientist | Carvi (Aug 2020 – Mar 2021)

Designed and implemented a serverless data pipeline to process high-frequency sensor data (GPS, Accelerometer, Gyroscope) from vehicle IoT devices. Developed a 'Safety Score' algorithm to quantify driving behavior and predict accident risks in real-time.

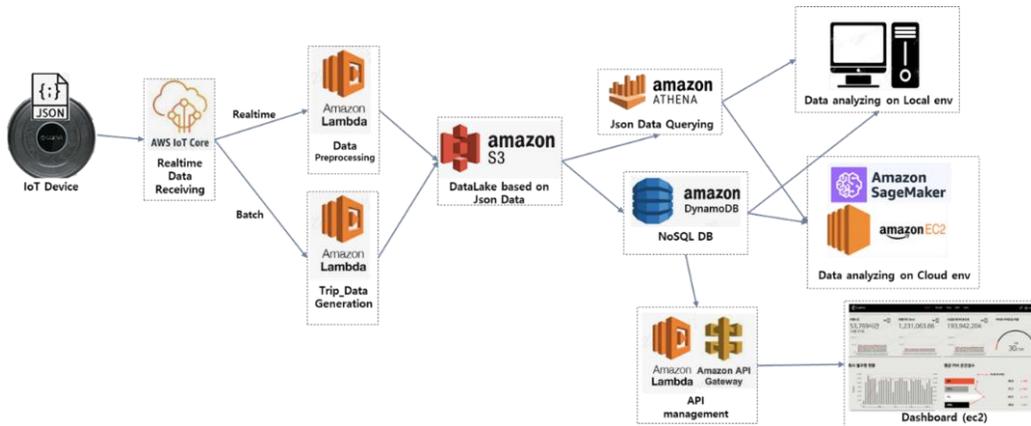


Figure: AWS Serverless Architecture for Real-time Vehicle Data Ingestion & Processing

Key Technologies: Python, AWS (Lambda, Athena, DynamoDB, S3), Scikit-learn

2. Automated ETL & Bond Asset Allocation Logic (GXIG ETF)

Data Scientist (Project) | Wealthspot (Dec 2024 – Current)

Engineered a centralized data warehouse on Snowflake, unifying fragmented financial data from Bloomberg and FactSet. Developed a hybrid investment logic combining Quantitative Factors with Deep Neural Networks (DNN) for the GXIG ETF launch.

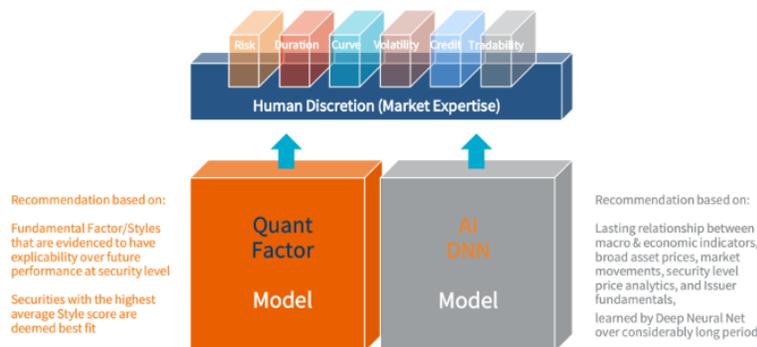


Figure: Asset Allocation Logic Flow & Data Integration Architecture

Key Technologies: Python, Snowflake, Bloomberg API, Deep Neural Networks (DNN), Quant Analysis

3. Enterprise LLM Pipeline for Financial Intelligence

Senior Data Scientist | Mirae Asset Securities (Mar 2021 – Dec 2024)

Architected an on-premise LLM solution using Llama3 and Qwen2 to translate and summarize real-time US market news. Built a secure RAG (Retrieval-Augmented Generation) system for internal customer service automation.

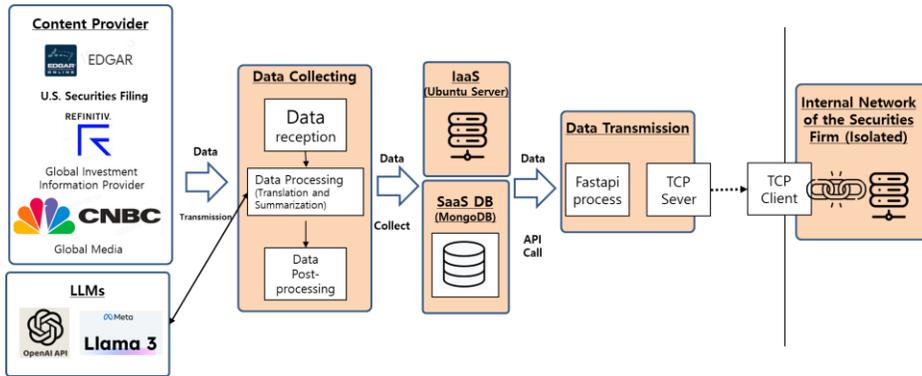


Figure: On-premise LLM Application Architecture & Data Flow

Key Technologies: Python, Llama3, LangChain, RAG, FastAPI, Docker, MongoDB

4. Semiconductor Manufacturing Optimization System

Data Scientist | Samsung Electronics (Jan 2015 – Aug 2018)

Analyzed massive wafer fabrication logs to identify bottlenecks in Photo/Etch processes. Developed fault detection algorithms improving defect detection accuracy by 30%.

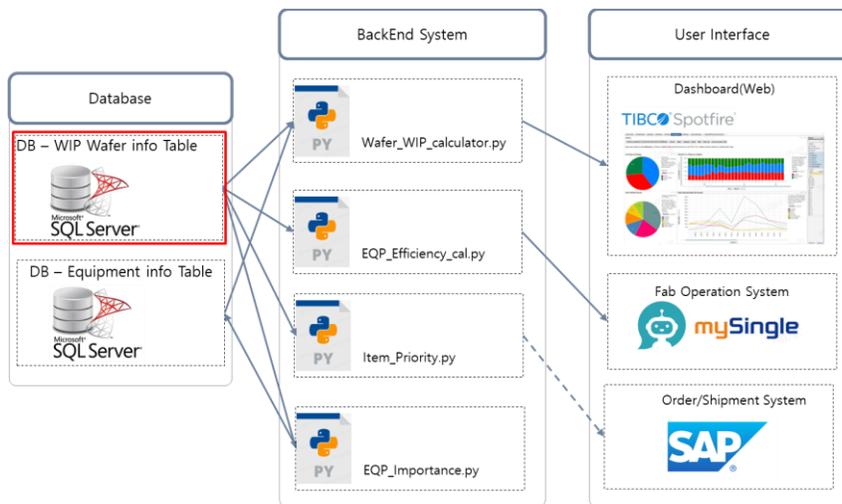


Figure: Smart Factory Data Pipeline & Scheduling Optimization System

Key Technologies: Python, MSSQL, Spotfire, Statistical Analysis

5. Research: Quantitative Text Mining & Sentiment Analysis

Master's Thesis | aSSIST (2017)

Investigated the relationship between quantitative text indicators (typos, special characters) and sentiment. Proved that incorporating these features significantly improves sentiment prediction accuracy in regression models.

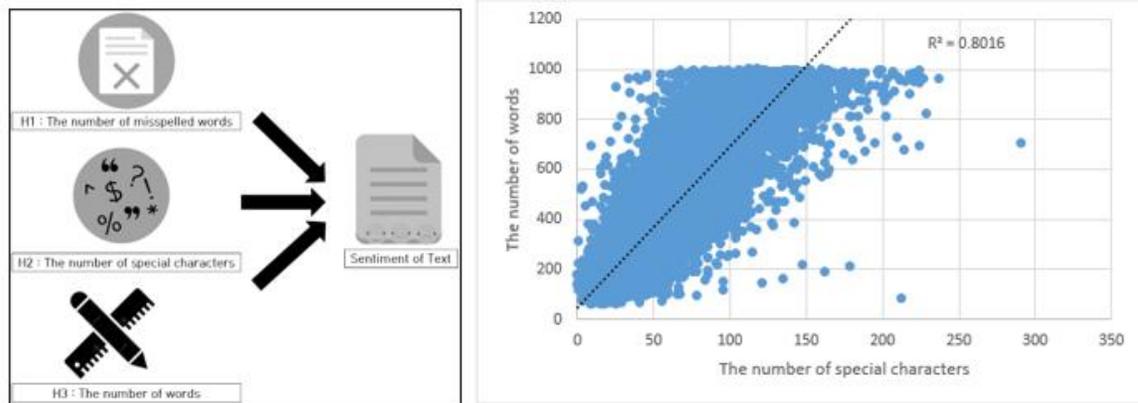


Figure: Correlation Analysis between Quantitative Indicators and Text Sentiment

Key Technologies: R, Python, Regression Analysis, NLP