

WONHA SHIN

Phone: +1) 551-245-2356 | Email: wshin7@ur.rochester.edu | Portfolio: leahnote01.github.io/ | LinkedIn: linkedin.com/in/wshin7

EDUCATION

University of Rochester

Master of Data Science GPA 3.43 / 4.0

Rochester, New York

Aug 2023 – Dec 2024

Pusan National University

Bachelor of Arts in International Business and Economics GPA 3.73 / 4.0 (Cum Laude)

Busan, South Korea

Mar 2011 – Aug 2018

PROFESSIONAL EXPERIENCE

VOESH New York

Administrative Accounting Specialist, Admin/Accounting

South Plainfield, New Jersey

Oct 2018 – Mar 2023

- Built a predictive time-series framework in Oracle DB, boosting sales forecasting accuracy by **15%** across 10K+ clients. Performed sales trend modeling to drive strategic planning and inventory optimization. Partnered with internal and external network cross-functional teams to streamline reporting and support data-driven decisions.

EASTERN AMERICAN CDC (Non-Profit Federal Community Bank)

Credit Analyst Intern, Credit Assessment

Englewood, New Jersey

May 2018 – Oct 2018

- Conducted quantitative credit analysis for 100+ SBA 504 loan applications, applying statistical and financial modeling techniques to assess repayment capacity and risk profiles. Performed trend and variance analysis on loan performance data, helping identify high-risk applicants and support risk-adjusted funding decisions.

PROJECTS & SKILLS

- **LLM-Based Real-Time Translation System with Kafka and MLOps Automation:** Developed and deployed a real-time English–Spanish translation system using Kafka (streaming), FastAPI (microservices), and Hugging Face Transformers, achieving ≤ 0.001 s latency under 10K+ concurrent messages. Implemented Prometheus and Grafana dashboards for system health monitoring and Weights & Biases (W&B) for drift detection, retraining, and model versioning. Validated scalability with load testing and optimized for both cost-efficient (200s interval) and enterprise-scale (ms-level frequency, 86K+/day) throughput.
- **200 Days Learning Posting Challenge ('24–'25):** Completed a 200-day study challenge focused on Deep Learning and MLOps, covering regression, PCA, decision trees, CNNs, transformers, and LLMs. Documented daily learning on GitHub and LinkedIn, building a knowledge base and engaging the ML community with technical write-ups.
- **E-Cigarette Perception Capstone Project (Chief Operations Manager):** Led NLP analysis of 100K+ multilingual social media posts using sentiment analysis and LDA topic modeling to identify health risks, misinformation, and demographic trends. Coordinated team operations, applied MLOps principles for reproducibility, and produced insights to support public health initiatives, preparing results for publication.
- **Microsoft Azure MLOps Pipeline Implementation:** Built a CI/CD pipeline on Azure using GitHub Actions, Docker, ACR, and ACI to automate model build, test, and deployment. Integrated with Azure ML for training, versioning, and continuous delivery, ensuring secure, production-ready workflows. Strengthened monitoring of KPIs for deployment reliability and scalability.
- **Real-Time Tweet Sentiment Analysis Pipeline :** Developed a streaming pipeline with PySpark, Databricks, and AWS to classify tweets in real time and store results in Delta Lake. Implemented structured streaming with windowed triggers to ensure low-latency, fault-tolerant sentiment analysis across large datasets.
- **Cross-Cultural NLP Analysis of Luxury Hotel Reviews in Europe:** Conducted topic modeling of 515K+ hotel reviews with LDA and scikit-learn pipelines to uncover cultural differences in traveler experiences. Generated insights for the hospitality and tourism industry, supporting data-driven market strategy.
- **Skills:** Infrastructure & Automation: Ansible (server provisioning & configuration), Docker, Kubernetes, GitHub Actions, Data Engineering & Streaming: Apache Kafka, PySpark, Databricks, Delta Lake, Airflow, MLOps & Monitoring: Azure ML, MLflow, Prometheus, Grafana, Weights & Biases (W&B), CI/CD pipelines, ML Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, Keras, Cloud Platforms: AWS (S3, EC2, Lambda, ECR), Microsoft Azure, Databases & Data Management: PostgreSQL, SQL Analytics, ETL pipeline development, Data Cleaning & Integration, Programming & Tools: Python (NumPy, Pandas, FastAPI), R (RStudio), Git (Version Control)