

# WONHA SHIN

Email: [wshin7@ur.rochester.edu](mailto:wshin7@ur.rochester.edu) | Project Portfolio: <https://leahnote01.github.io/>

## 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

- Designed and deployed a predictive time-series framework integrated with Oracle DB, improving sales forecasting accuracy by 15% across 10,000+ clients. Conducted time-series data analysis and sales trend modeling, enabling data-driven strategic planning and inventory optimization. Collaborated with cross-functional teams (sales, finance, and operations) to translate data insights into business strategies, supporting growth and scalability.

### 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:** Spearheaded end-to-end real-time multilingual translation system as a solo architect & engineer, integrating **Kafka** (streaming), **FastAPI** (async microservices), and **Hugging Face Transformers** for low-latency neural machine translation. Implemented server provisioning via **Ansible**, ensuring reproducible, scalable multi-node deployments. Integrated **Prometheus & Grafana** dashboards for system health monitoring and **Weights & Biases** for drift detection, retraining, and model versioning. Achieved  $\leq 0.001$ s translation latency and validated scalability through load testing of 10k+ concurrent messages.
- 200 Days Learning Posting Challenge ('24-'25):** Completed a 200-day deep learning and MLOps-focused study challenge, mastering foundational and advanced topics in ML, AI, and Deep Learning—from regression, PCA, decision trees, and SVMs to CNNs, transformers, and LLMs. Explored advanced themes including efficient transformers, bias mitigation, AI ethics, conformal prediction, and MLOps practices such as reproducibility, monitoring, and deployment. - Documented daily progress with technical write-ups on GitHub and LinkedIn, engaging the broader AI/ML community.
- E-Cigarette Perception Capstone Project (Chief Operations Manager):** Directed a full-cycle NLP project analyzing 100,000+ multilingual social media posts to uncover public perceptions of e-cigarette use and support public health initiatives. Led team coordination, modeling, MLOps planning, and data governance, while managing human annotation workflows and ensuring labeling quality. Applied sentiment analysis and LDA topic modeling to identify health risks, misinformation, and demographic trends. Preparing findings for publication focused on health discourses.
- Real-Time Tweet Sentiment Analysis Pipeline :** Built a real-time sentiment analysis pipeline with real-world using PySpark, Databricks, and AWS, streaming tweets from S3, classifying them with a pretrained transformer, and storing results in Delta Lake. Implemented structured streaming with windowed triggers for reliable, low-latency processing.
- Data Mining: Cross-Cultural NLP Analysis of Luxury Hotel Reviews in Europe- LDA Topic Modeling :** Led analysis of 515,000+ hotel reviews using LDA topic modeling to reveal cultural differences in traveler experiences across Europe. Leveraged scikit-learn pipelines to uncover trends, improving market insights for hospitality and tourism analysis.
- Skills:** Infrastructure & Automation: Ansible (server provisioning & configuration), Docker, Kubernetes, GitHub Actions, Terraform (basic), Data Engineering & Streaming: Apache Kafka, PySpark, Databricks, Delta Lake, Airflow MLOps & Monitoring: MLflow, Prometheus, Grafana, Weights & Biases (experiment tracking, drift detection, model versioning), CI/CD pipelines, ML & AI Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, Keras, Clouds: AWS (S3, EC2, Lambda, ECR), Databases & Data Management: PostgreSQL, SQL Analytics, ETL pipeline development, data cleaning & integration, Python (NumPy, Pandas, FastAPI), R (RStudio), Git (Version Control)