WONHA SHIN

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

Machine Learning Engineer with strong foundations in applied ML, data engineering, and scalable infrastructure. Experienced in designing, training, and deploying models with focus on reliability, performance, and automation. Skilled in Python, PyTorch, and cloud-native pipelines (AWS, Azure, Databricks) with production-grade MLOps practices (CI/CD, containerization, monitoring, drift detection).

SKILLS & PROJECTS

Programming & Tools: Python (NumPy, Pandas, FastAPI, Matplotlib, Seaborn), R (RStudio), Git, Shell Scripting
Data Engineering & Streaming: PySpark, Databricks, Delta Lake, Apache Kafka, Apache Airflow, Spark Streaming
Cloud Platforms: AWS (S3, EC2, ECR, Lambda, IAM), Azure (ACR, ACI, Azure ML), GCP (BigQuery, Vertex AI)
Databases & Data Management: SQL, PostgreSQL, MySQL, Data Warehousing, ETL Pipelines, Data Integration
Machine Learning & NLP: PyTorch, TensorFlow, Keras, scikit-learn, XGBoost, Hugging Face Transformers, LLMs
MLOps & Deployment: MLflow, Azure ML, W&B, Docker, CI/CD (GitHub Actions, GitLab CI), Prometheus, Grafana

- **LLM-Based Real-Time Translation System with MLOps Automation** | *Kafka • FastAPI • Hugging Face • W&B*:
Built a **low-latency real-time translation system** using Kafka and FastAPI microservices with Hugging Face Transformers in PyTorch, processing 10K+ concurrent requests. Deployed **containerized models** with Docker and automated CI/CD for scalable inference and version control. Integrated **Prometheus/Grafana** for monitoring and **W&B** for experiment tracking, drift detection, and retraining automation.
- **Real-Time Tweet Sentiment Analysis Pipeline** | *Databricks • PySpark • AWS • Delta Lake*
Developed a **structured-streaming pipeline** on AWS using PySpark to classify and monitor live tweet sentiment in real time. Designed **Bronze/Silver/Gold Delta architecture** for scalable ETL and optimized partitioning for **low-latency, fault-tolerant processing**. Created KPI dashboards and model performance tracking using Databricks.
- **Microsoft Azure MLOps Pipeline Implementation** | *Azure ML • Docker • GitHub Actions • ACR/ACI*
Built a **CI/CD pipeline** using Docker and GitHub Actions for automated model build, test, and deployment on Azure ML. Integrated Azure Container Registry (ACR) and Azure Container Instances (ACI) for containerized workflow.
- **E-Cigarette Perception Analysis (Capstone)** | *Python • NLP • scikit-learn • SQL • Tableau*
Led **NLP analysis** on 100K+ multilingual social posts to identify misinformation and health-risk patterns. Applied **sentiment analysis** and **LDA topic modeling** to uncover demographic-specific insights for public health strategies.
- **Cross-Cultural NLP Analysis of European Hotel Reviews** | *Python • LDA • Data Visualization*
Processed 500K+ multilingual reviews to identify **sentiment and topic trends** across European regions. Applied **TF-IDF vectorization** and topic modeling to reveal cultural patterns in customer feedback.
- **200 Days Learning & Knowledge-Sharing Challenge (2024–2025)**
Completed a 200-day continuous learning initiative focused on Data Analytics, Statistics, ML and DL foundations. Documented key insights and visualizations on GitHub and LinkedIn, building a public knowledge repository

PROFESSIONAL EXPERIENCE

VOESH New York   **South Plainfield, New Jersey**
Administrative Accounting Specialist, Admin/Accounting   *Oct 2018 – Mar 2023*
Built a predictive sales model in Oracle SQL, improving forecasting accuracy by 15% for 10K+ clients. Created analytical reports for executive decision-making, integrating data from finance and sales systems.

EASTERN AMERICAN CDC (Non-Profit Federal Community Bank)   **Englewood, New Jersey**
Credit Analyst Intern, Credit Assessment   *May 2018 – Oct 2018*
Conducted quantitative analysis for 100+ SBA loan applications, identifying risk and performance trends. Developed variance and trend analyses to optimize portfolio risk assessment and loan strategy.

EDUCATION

University of Rochester   **Rochester, New York**
Master of Data Science GPA 3.43 / 4.0   *Aug 2023 – Dec 2024*

Pusan National University   **Busan, South Korea**
Bachelor of Arts in International Business and Economics GPA 3.73 / 4.0 **(Cum Laude)**   *Mar 2011 – Aug 2018*