

Nishtha Wagh

nishtha.wagh.1999@gmail.com | +16232919222 | linkedin.com/in/nishthawagh99

Professional Summary

Senior Data Scientist with 4+ years building and operating production ML and LLM systems, from ambiguous problem definition to low-latency, monitored deployments driving measurable business impact.

Work Authorization: F-1 OPT (current) + STEM OPT valid to 2028 (no sponsorship required short term); Open to relocation

Technologies

ML Engineering: Model Deployment, Real-Time Scoring, Batch Pipelines, Feature Stores, Model Monitoring, Drift Detection

Machine Learning: XGBoost, Gradient Boosting, Logistic Regression, Propensity Models, Segmentation, NLP (BERT, Topic Modeling)

Programming: Python, SQL, PySpark, R

Cloud & Infra: AWS (SageMaker, Redshift, S3), GCP (BigQuery, Cloud Run, Vertex AI), Docker, Jenkins, CI/CD

Data Engineering: ETL Pipelines, Distributed Processing, Data Quality, Incremental Pipelines, Redshift Optimization

LLM Systems: Prompt pipelines, embeddings, text classification, evaluation workflows

Work Experience

Machine Learning Engineer, ASU Enterprise Partners, Arizona, United States

Mar 2024 – Present

- Designed and deployed a **containerized LLM inference pipeline on Cloud Run** with autoscaling, processing **20K+ weekly text inputs** under latency and cost constraints.
- Built **LLM-based sentiment and narrative classification workflows** using prompt structuring, batching, **embedding-based representations**, and post-processing for production reliability.
- Implemented a **BERT-based topic modeling system** for **2,200+ narratives**, reducing manual categorization from weeks to hours.
- Developed **segmentation models** using behavioral and demographic features to support downstream personalization systems.
- Built **production-grade BigQuery pipelines (SQL)** for feature generation, batch scoring, and monitoring with clear separation of offline and online workloads.
- Implemented **CI/CD workflows (Jenkins)** for model deployment, versioning, rollback safety, and environment parity across releases.

Senior Machine Learning Engineer, Amazon India (via Axio), India

May 2022 – June 2023

- Built and deployed **real-time risk scoring services on AWS SageMaker**, serving **500K+ credit decisions/day** in production.
- Engineered **200+ bureau and behavioral features using SQL and PySpark on Redshift**, integrating them into scalable pipelines.
- Designed **risk-based segmentation system** for credit line optimization, enabling automated decisioning.
- Built a **model monitoring framework** tracking drift across **200+ features**, with automated alerts and PSI-based thresholds integrated into production pipelines.
- Refactored **Redshift pipelines** using **incremental ETL patterns**, reducing runtime by **95% (8 hours to 25 minutes)** and enabling daily retraining.
- Led data quality investigations across **5M+ records**, improving model reliability by **12%**.

Machine Learning Engineer, Amazon India (via Axio), India

Jan 2021 – May 2022

- Deployed **production ML models** (Gradient Boosting, Logistic Regression) on AWS SageMaker supporting **2M+ BNPL users**.
- Built **batch scoring pipelines (Python, SQL)** processing **300K+ daily evaluations**, reducing manual underwriting effort.
- Engineered **80+ credit and behavioral features** from bureau data, transaction logs, and repayment history.
- Developed **automated evaluation and monitoring pipelines** for model performance and portfolio health, supporting continuous production oversight.
- Built **Looker dashboards** for model outputs, latency metrics, and system reliability for Product, Risk, and Finance teams.

Education

Arizona State University, USA, Master of Science in Data Science; **GPA: 3.90**

Aug 2023 – May 2025

Nirma University, India, Bachelor of Technology in Computer Engineering; **GPA: 3.46**

Aug 2017 – Jun 2021