Nishtha Wagh

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

Professional Summary

Senior Data Scientist with 4+ years shipping production ML systems and owning end-to-end analytics, from ambiguous problem to deployed solution driving measurable business impact.

Technologies

Languages: Python, SQL, PySpark, R
Machine Learning: Logistic Regression, Random Forests, Gradient Boosting, XGBoost, Clustering, Feature Engineering
Cloud & Data Platforms: AWS (SageMaker, Redshift, S3), GCP (BigQuery, Cloud Run, Vertex AI), Azure
Experimentation & Statistics: A/B Testing, Hypothesis Testing, Causal Inference
BI & Visualization: Power BI, Tableau, Looker
Production & Tooling: Docker, CI/CD, Model Monitoring, Git

Work Experience

DATA SCIENTIST, Enterprise Partners, Arizona, United States   Mar 2024 – Present

- Built **clustering models (K-means, hierarchical)** on behavioral and demographic data to segment **50K+ alumni**, identifying underperforming cohorts that informed annual personalization strategy.
- Designed and analyzed **A/B experiments** testing personalized content, improving CTR and registrations **(7% lift)** with zero unsubscribe impact.
- Deployed production **LLM pipeline** on **GCP Cloud Run** processing **20K+ weekly narratives**, extracting sentiment and interest signals that reduced qualitative analysis from **weeks to hours**.
- Built BERT-based topic models on 2,200+ alumni stories, surfacing affinity patterns that reduced manual categorization.
- Built **Power BI dashboards** linking engagement metrics to donation conversion rates, enabling alumni teams to optimize personalization and outreach strategies.

SENIOR DATA SCIENTIST, Amazon India (via Axio), India   May 2022 – June 2023

- Built and deployed **gradient boosting risk models** on **AWS SageMaker** using **200+ bureau and behavioral features**, serving 500K+ daily scoring requests.
- Designed credit-line optimization using **risk-based segmentation,** enabling portfolio growth and maintaining default targets.
- Engineered behavioral features from transaction patterns using **SQL and PySpark**, improving cross-sell propensity model performance across customer cohorts.
- Led root-cause analysis on bureau score anomalies across **5M+ customer records**; findings drove data pipeline fixes that improved scoring accuracy by **12%**, presented to **VP of Risk**.
- Refactored feature pipelines in **Redshift** using **incremental ETL patterns** orchestrated via **Jenkins**, reducing computation time by ~**95% (8 hours to 25 minutes)** and enabling daily model retraining.
- Implemented **Python-based model monitoring** tracking drift across **200+ features** to detect population shifts and prevent performance degradation.

DATA SCIENTIST, Amazon India (via Axio), India   Jan 2021 – Apr 2022

- Built and deployed **risk models (Gradient Boosting, Logistic Regression)** on **AWS SageMaker** for Amazon BNPL serving **2M+ customers (production AUC 0.78).**
- Designed **controlled experiments** on cross-sell campaigns using **Python and SQL**, optimizing targeting strategies through propensity modeling.
- Engineered **80+ credit and behavioral features** in **Redshift** from bureau data, transaction logs, and repayment history.
- Built **batch pipelines** processing **300K+ daily evaluations**, reducing manual underwriting requirements.
- Built **Looker dashboards** tracking model outputs and portfolio health for **20+ stakeholders** across Product, Risk, & Finance.

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