

Jonesh Shrestha

Healthcare Data Scientist | Claims Analytics | Applied Machine Learning

Chicago, IL | (312) 973-9061 | contact@joneshshrestha.com | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

PROFILE

Healthcare data scientist with 2+ years building SQL/ODI pipelines for U.S. payer claims, eligibility, and pharmacy data, plus graduate AI research and projects in provider risk modeling, anomaly detection, and model evaluation.

TECHNICAL SKILLS

Programming & Libraries: Python, SQL, pandas, NumPy, scikit-learn, PyTorch, TensorFlow

Machine Learning & Evaluation: Gradient Boosting (XGBoost), Random Forest, Isolation Forest, SMOTE; feature engineering, anomaly detection, threshold optimization, precision, recall, ROC-AUC

Healthcare Data: U.S. payer claims, eligibility, pharmacy; ICD-10, CPT/HCPCS, DRG, NDC, claims adjudication, HIPAA/PHI

Data Engineering & Cloud: PL/SQL, Oracle Data Integrator (ODI), ETL, data profiling, AWS (S3, EMR, Athena, SageMaker), Snowflake, PySpark/Spark, Git, Docker

EXPERIENCE

DePaul University | AI Research Assistant

Chicago, IL | Jul 2025 - Present

- Benchmarked PaddleOCR/PP-OCRv5 and MMRotate for scientific-chart text detection on CHART-Info (8,343 training and 3,280 test charts); a Swin-Base RoI Transformer achieved **0.8957 total detection IoU**.
- Built a QA-controlled conversion pipeline for 140,669 block annotations, producing **170,184 word-level labels** with a **94.4% conversion acceptance rate**; used IoU and normalized character error rate for error analysis.
- Developing rotation-aware grouping to reconstruct predicted words into semantic text blocks for chart understanding.

Cotiviti Nepal Pvt. Ltd. | Software Engineer

Kathmandu, Nepal | Jul 2021 - Jul 2023

- Built SQL, PL/SQL, and Oracle Data Integrator (ODI) pipelines to standardize HIPAA-regulated claims, eligibility, and pharmacy data for **15+ U.S. payer clients**, working directly with PHI, ICD-10, CPT/HCPCS, DRG, NDC, and adjudication fields.
- Translated client and vendor layouts into Cotiviti's Common Data Format; authored import specifications and custom mappings and ran manual and automated Data Profiler checks to validate critical fields.
- Tested client and payer integrations; resolved data and mapping defects with developers, QA, and project managers through production handoff; trained **10+ engineers** on ODI workflows.

Kathmandu University | Graduate Research Assistant (Part-time)

Dhulikhel, Nepal | Mar 2021 - Apr 2022

- Built CNN and Transformer-based Nepali image-captioning models on Flickr8k for low-resource NLP; curated captions and evaluated outputs with BLEU and METEOR.

SELECTED PROJECTS

U.S. Healthcare Fraud, Waste & Abuse Detection | Provider Risk Modeling

Oct 2025 - Nov 2025

- Engineered 47 provider-level features from **558,211 synthetic Medicare-inspired claims** and 138,556 beneficiary records across 5,410 providers (9.35% positive).
- Using training-only SMOTE, Random Forest achieved **0.9546 ROC-AUC** on 1,082 test providers after benchmarking XGBoost and Isolation Forest; a **validation-tuned 0.46 threshold** yielded **53.90% precision and 82.18% recall**.

Yelp Hybrid Recommendation System | Big Data ML Pipeline

Sep 2025 - Nov 2025

- Built an AWS EMR/PySpark recommender on 1M+ Yelp reviews; stacking Spark ALS with a random forest over content and sentiment features reduced test RMSE **8% (1.55 to 1.43)** and MAE **16% (1.25 to 1.05)**.

EDUCATION

DePaul University | M.S. Computer Science, AI Concentration | GPA: 3.9/4.0

Jun 2026

Kathmandu University | B.S. Computer Science

Jun 2020

PUBLICATION & HONORS

Toward Production-Ready Federated Learning in Healthcare: Privacy, Orchestration, and Governance in MLOps

Honors: Graduate Presidential Scholarship | Graduate Research Assistant Program Award | Upsilon Pi Epsilon