APAAR SAROJ

+1 (602) 616-5991 | apaarsaroj@gmail.com | linkedin.com/in/apaarsaroj | github.com/apaarsaroj
New York, NY

Summary

3+ years of experience in analytics, machine learning and artificial intelligence. Recent ASU MS in Information Technology graduate (4.0 GPA). Targeting AI/Data Science/Data & ML Engineering roles to build end-to-end pipelines, Gen-AI products, and bring systems thinking to problems that actually matter

Capstone: Built end-to-end ML pipeline scoring healthcare financing efficiency across 52 countries using Shannon entropy weighting, panel regression, K-Means clustering, SHAP-explained Random Forest deployed as a live Streamlit dashboard

Technical Skills

Languages & Tools: Python (Pandas, NumPy, Scikit-learn), SQL, R (basic), Hadoop, HiveQL, Docker, Git, jupyter, ETL
Machine Learning & Stats: Random Forest, K-Means Clustering, Panel Regression (Fixed Effects), Feature Engineering, SHAP, Model Evaluation, Baseline Evaluation, Shannon Entropy, Cross-Validation (GroupKFold)
NLP & Deep Learning: BERT, PyTorch, Transformers, Transfer Learning
Visualization: Streamlit, Tableau, Power BI, Matplotlib, Seaborn
Cloud: AWS (foundational), Model Deployment, Containerization

Professional Experience

Business Analyst · EXL · Big Six UK Energy Provider *Sep 2022 – Jul 2024*

- Built and ran 20+ daily data engineering scripts processing 100K+ financial transactions for cash matching, unallocated transactions, and balance reconciliation across a regulated UK energy portfolio serving ~10M households
- Created 6+ Power BI dashboards for regulatory compliance, cash flow tracking, and final credits reporting. Used by Finance leadership for strategic planning and audit readiness. Recognized by client for high-quality, on-time delivery in fast-paced, high-pressure situations
- Wrote and automated recurring financial reports using SQL and Python, cutting turnaround time from 24 hours to under 1 hour. Took ownership of new processes from the technical leadership and streamlined reporting workflows

Data & BI Developer · Team Computers · Wind Energy Sector *Jul 2021 – Sep 2022*

- Built python predictive models using wind turbine sensor data to identify downtime drivers, cutting unplanned turbine downtime 20% and maintenance costs 15%
- Integrated weather forecast data using SQL and python, to build analytical energy production models improving resource planning accuracy and operational efficiency by 15%
- Built real-time Tableau dashboards tracking turbine KPIs (vibration, temperature, power output) for operations teams, enabling faster anomaly detection and reducing mean time to resolution

Projects

Healthcare System Efficiency Evaluation Framework · ASU *Jan 2026 to May 2026*

GitHub: github.com/apaarsaroj/healthcare-efficiency-framework | Live: healthcare-efficiency-framework.streamlit.app

- Engineered a composite efficiency score for 52 countries, using Shannon entropy to derive objective variable weights from 13 indicators across WHO, OECD, and World Bank.
- Modeled drivers with fixed-effects regression and SHAP-explained Random Forests, revealing financial protection over spending as the top coverage driver, and shipped a Dockerized Streamlit dashboard with policy simulator.
- Stack: Python, scikit-learn, statsmodels, SHAP, Streamlit, Pandas, NumPy, Docker

Cross-Domain Tone Classification · ASU *Fall 2025*

GitHub: github.com/apaarsaroj/Cross-Domain-Tone-Classification

- Studied whether you can train **BERT models** on incompatible label systems (one dataset labeled by tone, another by emotions) and still map them to a shared target using a small calibration set.
- The smaller, more aligned dataset (2.6K samples) hit 54.6% accuracy on 3-class tone classification. The larger but misaligned dataset (8K samples) only got 46.5%. Showed that domain fit matters more than size in transfer settings.
- Stack: BERT, PyTorch, Transformers, Python

Education

M.S. Information Technology · Arizona State University, Tempe AZ *GPA: 4.00 / 4.00 | 2024 – 2026*

Relevant Coursework: Natural Language Processing (A+), Advanced Big Data Analytics & AI (A), Statistical Foundations for IT (A+), Advanced Database Management Systems (A), Data Visualization & Reporting (A), Information Systems Security (A)

B.Tech Computer Science · Kurukshetra University, India *2015 – 2019*