Anav Lamba

+91 9650570773 | anavlamba94@gmail.com | github.com/anav94 | anavlamba-portfolio.vercel.app

Education

Manipal University Jaipur — Jaipur, India
Bachelor of Technology in Computer and Communication Engineering; CGPA: 9.44 — *2022 – 2026*
- **Relevant Coursework:** Design and Analysis of Algorithms, Computer Networks, Object Oriented Programming, Foundations of Data Science, Relational Database Management Systems, Artificial Intelligence and Machine Learning

Springdales School, Pusa Road — New Delhi, India
CBSE Class XII; Marks: 92% — *2008 – 2022*

Experience

Project Intern — May 2025 – August 2025
BPAAS Solutions — *Remote*
- Automated LinkedIn/web capture and Excel email reports with Selenium (Python), cutting manual effort by 50–60% (4–6 hrs/week) and reducing errors to 1–2% via validations
- Scheduled headless runs (GitHub Actions/cron) to move refresh from weekly to daily; standardized CSV/XLS outputs used by sales/ops in under 1 hour
- Added structured logging and failure alerts; maintained 99% run success with 6-minute average runtime per workflow, enabling same-day outreach on new leads

Projects

CrisisPulse — Real-time Event Intelligence [GitHub | Live] 2025
Python (FastAPI, asyncio), Kafka/Redpanda, Postgres, ClickHouse, Streamlit, Docker, Prometheus/Grafana
- Real-time pipeline ingesting hazard feeds (USGS) + synthetic events, deduplicating, geo-enriching (H3), and serving a live map & API
- p95 latency ˜3s end-to-end; ClickHouse queries ¡1s; SLA ¡5s for sev 0.8
- Dropped ˜70% duplicates in a 7.5k-event burst via MinHash/LSH; sustained ˜1,000 msgs/min throughput on M2
- Production touches: exposed /health and versioned /openapi.json; Prometheus /metrics with alerts (p95¿5s, ingest gap¿60s); Grafana dashboards
- Data quality & drift: Great Expectations checks on schema/nulls/ranges; Evidently drift report gating loads and raising warnings
- Ops & CI/CD: GitHub Actions (lint/tests/badges), scheduled backfills/compaction via cron, and roll-forward playbook for failed ingests

ChurnWatch – Explainable Churn Prediction [GitHub | Live] 2025
Python, scikit-learn, XGBoost, MLflow, SHAP, FastAPI, Docker
- Built an end-to-end churn prediction system on telecom data using scikit-learn and XGBoost (ROC-AUC 0.85)
- Tracked experiments and artifacts with MLflow; integrated SHAP for customer-level explanations and an eligibility playbook of retention actions
- Served a FastAPI scoring API (Docker-ready) and delivered a Streamlit insights app with risk cohorts and what-if analysis
- Generated Evidently reports for data/performance drift monitoring

CommerceStack – Analysis Stack [GitHub | Live] 2024
dbt, DuckDB, Streamlit, Python, SQL
- Built a decision-ready analytics stack on 100k+ Brazilian e-commerce orders using dbt, DuckDB, and Streamlit
- Modeled raw Kaggle data into staging/marts with fact and dimension tables for weekly revenue, retention, and cohort KPIs
- Automated ELT builds with dbt seeds and models on deploy, scaffolded data quality checks with Great Expectations
- Authored SQL/dbt transformations with window functions for order and payment insights
- Delivered a Streamlit dashboard with KPIs, weekly revenue trend

Technical Skills

Languages: Python, C/C++, Java, JavaScript, SQL
Web Technologies: HTML, CSS, ReactJS, REST API Integration, FastAPI
Databases: MySQL, MongoDB, DuckDB, Excel
Tools & Platforms: Git & GitHub, AWS (EC2, S3, Lambda), Selenium Web Automation, Docker, MLflow
Data Science & ML: scikit-learn, XGBoost, SHAP, pandas, NumPy, dbt, Streamlit, Great Expectations, Evidently
Networking & Observability: TCP/IP, HTTP, DNS fundamentals, Kafka basics, NetFlow/IPFIX concepts, OpenTelemetry (metrics/logs/traces) basics

Certifications

IBM: Introduction to Software Engineering
CISCO: CCNAv7: Introduction to Networks, Switching, Routing, and Wireless Essentials; Python Essentials 1 & 2
NPTEL: Design and Analysis of Algorithms; Database Management Systems (Relational DBMS)
Honors: College Dean's List (5x)