
SUMMARY

Computer Science graduate with two data analytics internships and hands-on experience building analytics solutions using Python and SQL. Worked on projects spanning geospatial monitoring, review authenticity, and financial analytics, covering data preparation, machine learning, and interactive dashboards.

TECHNICAL SKILLS

Languages: Python, SQL, R

Data Analysis: Pandas, NumPy, Excel

Machine Learning: Scikit-learn, PyTorch, Hugging Face Transformers, SHAP, Sentence-Transformers

Graph Analytics: NetworkX, Louvain Community Detection

Databases: PostgreSQL, MySQL, MongoDB, Supabase

Visualization & BI: Power BI, Plotly

Application Development: FastAPI, Streamlit

Geospatial Analytics: Google Earth Engine, Remote Sensing, LULC Change Detection

Tools: Git, Jupyter Notebook, OpenAI API

WORK EXPERIENCE

Research Data Analyst — MRSAC (Maharashtra Remote Sensing Application Centre)

Dec 2025 – Mar 2026

- Processed multi-temporal satellite imagery across 15+ Maharashtra districts on Google Earth Engine to map urban expansion over a 5-year window.
- Built an early-warning module flagging encroachment hotspots via LULC change-detection, cutting analyst review effort by surfacing only high-priority areas.
- Piped live AQI, thermal-anomaly, and UV data into an analytics backend; added a natural-language query layer (via OpenAI) so non-technical staff could query data in plain English.

Data Analyst Intern — Dealintra Infotech Pvt. Ltd.

May 2025 – Jun 2025

Review Authenticity Engine

- Processed 50,000+ e-commerce reviews through an NLP pipeline (preprocessing, TF-IDF extraction, model evaluation) to predict review authenticity.
- Built a review analytics pipeline that identified coordinated manipulation by combining review timing, seller relationships, and linguistic patterns with model predictions.
- Developed interactive dashboards surfacing authenticity scores, seller-credibility metrics, and suspicious activity for fraud investigation.

PROJECTS

APGuard — Accounts Payable Leakage Detection

Python · PostgreSQL · FastAPI · SQLAlchemy · Pandas · Streamlit · Plotly · pytest

- Built a rule-based analytics system to identify payment leakage and contract compliance issues across Accounts Payable transactions.
- Generated a synthetic dataset of 50,000+ invoices, purchase orders, contracts, goods receipts, and payments with realistic exceptions—including duplicate invoices, contract price violations, tax mismatches, and missing purchase orders—to evaluate rule performance.
- Developed 8 business validation rules covering duplicate payments, three-way match failures, contract price checks, invoice splitting, and tax validation, with automated pytest coverage for edge cases.
- Designed a relational PostgreSQL schema and ETL pipeline integrating six data sources with indexed joins; built a FastAPI backend and Streamlit dashboard for vendor drill-down, risk scoring, and CSV export.

SentraScope — Environmental Analytics

Google Earth Engine · PostgreSQL · React · TypeScript · Node.js/Express

[GitHub](#) | [Live Demo](#)

- Built an environmental analytics platform integrating live air quality (WAQI/CPCB), NASA FIRMS thermal-hotspot, OpenUV, and Earth Engine urbanization data into a unified monitoring dashboard.
- Designed a PostgreSQL schema (Drizzle ORM, Supabase) for geospatial time-series queries across historical trends and live anomaly alerts at district scale.
- Implemented JWT auth with Brevo email verification and a shared OpenAPI spec generating typed Zod validators, plus scheduled report emails and configurable alert thresholds.

Review Authenticity Engine — Fraud & Collusion Detection System

Python · PyTorch · Hugging Face Transformers · NetworkX · SHAP · Streamlit

[GitHub](#) | [Hugging Face Model](#)

- Built a two-layer fraud detection system: a fine-tuned DistilBERT classifier trained on Amazon Reviews with synthetic fakes at three sophistication tiers — 100% recall on template/combinatorial tiers, 99.3% on paraphrased fakes.
- Engineered a reviewer graph weighted by timing, text similarity, and product overlap; Louvain community detection surfaced 132 dense clusters, independently catching ~35% of fake reviewers on behavioral signal alone.
- Identified and fixed a train/test data leak (>95% overlap) inflating F1 to ~0.995; rebuilt the pipeline with a leak-safe split and deployed SHAP-based explanations for predictions.

EDUCATION

B.E. — Computer Science & Engineering 2022 – 2026

SB Jain Institute of Management & Research, Nagpur

Relevant coursework: Machine Learning, Database Systems, Data Warehousing, Computer Networks

12th (HSC) — Taywade College, Nagpur, 2022 · 10th (SSC) — Hadas High School, Nagpur, 2020

CERTIFICATIONS

-
- Ask Questions to Make Data-Driven Decisions (Coursera), Oct 2025
 - Big Data Computing — NPTEL (IIT), 78%