

Shivam Rajput

Data Science & Machine Learning | Statistics, NLP & Applied AI

Dehradun, Uttarakhand | +91 6398994844 | shivamrajput.datascientist@gmail.com

LinkedIn • GitHub • Portfolio • LeetCode

SUMMARY

Final-year B.Tech CSE (AI & ML) student with four end-to-end projects spanning large-scale analytics, statistical experimentation, machine learning, NLP, human-reviewed ML workflows, and RAG. Processed 15.95M records, evaluated a 588K-user experiment, and built tested applications using Python, SQL, scikit-learn, FastAPI, Docker, and MLflow.

EDUCATION

Uttaranchal University, Dehradun

2023-2027

B.Tech, Computer Science & Engineering (AI & ML)

CGPA: 8.7/10

Relevant Coursework: Statistics, Machine Learning, Data Structures & Algorithms, DBMS, OOP

TECHNICAL SKILLS

Programming & Data - Python, SQL, Pandas, NumPy, PyArrow, Microsoft SQL Server

Statistics & Machine Learning - A/B testing, confidence intervals, hypothesis testing, classification, regression, cross-validation, forecasting, class imbalance, model evaluation, SHAP

NLP, LLMs & RAG - TF-IDF, text classification, topic modeling, embeddings, BM25, CrossEncoder reranking, ChromaDB, LangGraph, human-in-the-loop workflows

Engineering & MLOps - FastAPI, React + Vite, Streamlit, Docker, MLflow, pytest, Ruff, Git, GitHub Actions

SELECTED PROJECTS

Customer Complaint Intelligence Platform

Jun 2026 • GitHub • Demo

Python, Pandas, PyArrow, Prophet, scikit-learn, Streamlit, Docker

- Engineered a chunked pipeline for 15.95M CFPB complaints, converting an 8-9 GB CSV into reusable Parquet outputs for local analytics.
- Built executive analytics, company-risk signals, NLP routing and forecasting; documented product/issue classifier limits and 14/14 core tests.
- Achieved 3.57% MAPE with a Prophet forecast on a documented six-month holdout.

Marketing A/B Testing and Statistical Experiment Analysis

Aug 2026 • GitHub

Python, Pandas, SciPy, Statsmodels, Matplotlib

- Evaluated 588,101 users and measured 2.55% advertisement conversion versus 1.79% for the PSA control group.
- Estimated +0.769 percentage-point uplift (43.1% relative; 95% CI: +0.595 to +0.943 points) and translated the evidence into a cost-aware rollout recommendation.

Agentic ML Audit Copilot

Jun-Jul 2026 • GitHub • Demo

Python, scikit-learn, LangGraph, MLflow, SHAP, FastAPI, Streamlit, Docker

- Designed a deterministic-first audit workflow for data-quality, target-leakage, class-imbalance and modeling risks before training.
- Implemented conditional routing and a human-review stop/resume gate, using the LLM only for grounded explanations, reports and Q&A.
- Added automated pytest coverage, baseline modeling, MLflow tracking, SHAP explanations, API serving and Docker packaging.

Enterprise RAG Assistant

Jun-Aug 2026 • GitHub • Demo

Python, FastAPI, React + Vite, ChromaDB, BGE, BM25, CrossEncoder, Pandas, Docker

- Built source-grounded Q&A across seven validated file formats, routing tabular questions to deterministic Pandas analysis and semantic questions to BGE + BM25 hybrid retrieval.
- Integrated deduplication, CrossEncoder reranking, citations and conversation context while preserving source metadata.
- Recorded a 0.947 composite on a 29-case strict hybrid RAGAS evaluation, including 0.966 faithfulness; the separate 1,642-case production benchmark still has an open latency acceptance gate.

CERTIFICATIONS & CODING

HackerRank: SQL (Intermediate), Jul 2026

LeetCode: 511 problems solved, including 242 Medium and 35 Hard