RAJVEER SINGH PALL

Jabalpur, Madhya Pradesh, India · rajveerpall04@gmail.com · +91 7000700355 · linkedin.com/in/rajveer-singh-pall-80b385290 · github.com/Rajveer-code · rajveer-code.github.io

Undergraduate CS researcher with nine papers spanning quantitative finance, causal econometrics, healthcare ML, and LLM fairness auditing — sole author on six, including a runtime LLM fairness auditor (arXiv) and a statistically-gated equity-prediction framework (under review, Quantitative Finance & Economics). Builds full ML systems end-to-end, from federated-learning pipelines to deployed RAG and voice-AI products. Seeking research and machine-learning internships.

EDUCATION

B.Tech, Computer Science & Business Systems — Gyan Ganga Institute of Technology and Sciences, Jabalpur *2023 – 2027*

CGPA: 7.8 / 10

Certificate Program in Data Science — IIT Guwahati *2024 – 2025*

Advanced ML, analytics, internship-based learning

RESEARCH PUBLICATIONS

Comprehensive Evaluation of ML for Type 2 Diabetes Risk Prediction: External Validation & Fairness Analysis

Co-author with S. Yadav (corresponding) *IEEE — Accepted*

- Externally validated an XGBoost diabetes-risk model trained on NHANES (n=15,685) against BRFSS (n=1.29M) — internal AUC 0.794 dropped to 0.717 externally (−9.7%); fairness analysis exposed a 0.135 AUC gap between elderly and young patients. TRIPOD-AI–compliant reporting with SHAP-based interpretability.

When the Gate Stays Closed: Near-Zero Cross-Sectional Predictability in Large-Cap NASDAQ Equities Using an IC-Gated ML Framework

Sole Author *Quantitative Finance & Economics (AIMS) — Under Review*

- Built a two-stage conjunctive statistical gate (HAC Newey–West t-test AND permutation test) to flag false-discovery risk in ML equity-prediction claims; across 12 walk-forward folds (1,512 OOS days) the gate correctly never fired (mean IC −0.0005), cutting an 11.8% naive false-deployment rate to 0.0%.

Privacy-Preserving Federated Learning for Diabetes Risk Prediction

Co-author with S. Yadav *Computer Methods & Programs in Biomedicine — Under Review*

- Built a federated learning pipeline (FedAvg / FedProx / FedNova / SCAFFOLD) across simulated clinical sites on NHANES (n=15,650), externally validated on BRFSS (n=1.28M) — external AUC 0.757, a 40% smaller generalization gap than centralized training, and calibration error (ECE) reduced from 0.276 to 0.001. PyTorch, Flower, Opacus.

Who Bears the Burden? Heterogeneous Treatment Effects in Mortgage Lending

Sole Author *J. Real Estate Finance & Economics — Submitted Jun 2026*

- Applied Causal Forest double machine learning to 42.3M HMDA mortgage applications (2020–2024); found a −9.39pp conditional Black-applicant denial penalty (t=−131.8) vs. −14.95pp unconditional, with manual underwriting widening the gap to −14.79pp. EconML, LightGBM, Polars at scale.

Persistent Racial Disparities in U.S. Mortgage Approval: Evidence from 42 Million Applications, 2020–2024

Sole Author *Journal of Housing Economics — Submitted Mar 2026*

- Companion study on the same HMDA panel using DiNardo-Fortin-Lemieux reweighting and within-lender fixed effects (74.6% of the gap is within-lender), a regression-discontinuity design at the 80% LTV threshold, and a difference-in-differences around the Fed's 2022 tightening cycle; partial-identification bounds show 44–55% of the gap remains unexplained under conservative assumptions.

Cross-Platform Fairness Failure in Mental-Health NLP Classifiers

Co-author with S. Yadav *IEEE Trans. Neural Networks & Learning Systems — Submitted Apr 2026*

- Ran a 5-axis fairness audit of transformer mental-health text classifiers (BERT/RoBERTa); performance collapsed off the platform they were trained on (AUC −30 to −39%, ECE up to 0.54, disparate impact <0.17) — a generalization failure invisible to single-platform evaluation.

ARIA: A Runtime 5-Axis Fairness Audit for Free-Form LLM Output

Sole Author *arXiv preprint (cs.LG/cs.CY) — Sep 2026*

- Built a runtime fairness auditor running inline counterfactual disparate-impact / equal-opportunity checks on free-form LLM output — catches 33–39% of fairness failures other guardrails miss, at ~1.2s latency, 6.5GB VRAM, and zero cloud calls. Open-sourced as the `aria-audit` PyPI package.

Overcoming the Transaction Cost Trap: Cross-Sectional Conviction Ranking in ML Equity Prediction
Sole Author *SSRN — Working Paper*

- Showed naive probability thresholds destroy ML equity alpha through turnover; reframed calibrated ensemble probabilities as ordinal conviction, holding only the top-ranked asset daily — IC 0.0197 (p=0.034, 1,512 days), 45.8% annualized return at Sharpe 1.18, break-even at 24.2 bps (4.8× typical institutional friction).

IndiaFinBench: A Benchmark for LLMs on Indian Financial Regulatory Text
Sole Author *Target: EMNLP 2026 — Draft*

- Built the first LLM benchmark for Indian financial regulatory text (SEBI + RBI filings) — 406 QA pairs over 192 documents, evaluated 12 LLMs (Gemini 2.5 Flash led at 89.7%), with a hybrid BM25+FAISS retrieval pipeline reaching Recall@5 of 0.785.

INDEPENDENT SYSTEMS & DEPLOYED PROJECTS

ARIA Assistant — Local Voice AI Agent
github.com/Rajveer-code/aria-assistant · Demo: aria-assistant.vercel.app

- Local voice-AI pipeline (faster-whisper → Qwen3-8B → 5-axis fairness audit → Piper TTS) with 16 callable tools, running end-to-end on a single RTX 4060 (≤7.1GB VRAM) with zero cloud LLM calls.

FinSight — Earnings-Call Intelligence Pipeline
github.com/Rajveer-code/Finsight

- FinBERT + RAG pipeline over 14,584 earnings-call transcripts across 601 companies; energy-sector signal reached an information coefficient (IC) of +0.31.

FinSight-Web — Interactive Research Monograph
finsight-web-rust.vercel.app

- Next.js 16 + React Three Fiber scroll-driven site presenting FinSight's findings as an interactive 3D monograph.

SereneSpace — Anonymous Student Mental-Health Hub
serenespace-e5ev.onrender.com · Smart India Hackathon 2025 — selected, 50 of 250 teams

- Built and deployed an anonymous peer-support hub for student mental health; sole developer on a 5-person team.

Earlier Hackathon Projects:

- Vaha — Automotive Predictive Maintenance (Techathon); sole developer, team of 5. vaha-project.onrender.com
- Beneficiary Credit Scoring System (SIH25145); sole developer, team of 5. credit-scoring-app-12lp.onrender.com

PROFESSIONAL EXPERIENCE

Data Science Intern — Unified Mentor Pvt. Ltd. *Apr – Jul 2025*

- Led end-to-end ML model development — data cleaning, feature engineering, and SHAP-based interpretability for client deliverables.

Python & Web Development Intern — InternPe *Sep – Oct 2024*

- Built backend Flask APIs with CRUD functionality and automation scripts for workflow optimization.

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript, Bash
Machine Learning: PyTorch, scikit-learn, XGBoost, LightGBM, CatBoost, SHAP
Causal Inference & Statistics: EconML (Causal Forests / DML), statsmodels, Polars
NLP & LLM Tooling: HuggingFace Transformers, FAISS, ChromaDB, LangChain, LangGraph, Ollama, Groq
Federated & Privacy-Preserving ML: Flower, Opacus (Differential Privacy)
Web & Deployment: FastAPI, Flask, Next.js, React, Three.js, Docker, Vercel, GCP

CERTIFICATIONS

Oracle Academy — SQL Programming, Database Design · Cisco Networking Academy — Cybersecurity Essentials, Introduction to Cybersecurity, Packet Tracer · Deloitte Data Analytics Job Simulation (Forage)