

Ankit Raj

 ankit-raj00 |  ankit-raj |  Portfolio |  ankit_r@hs.iitr.ac.in |  6204679511

EDUCATION

2027 Indian Institute of Technology, Roorkee

CGPA: 8.1

EXPERIENCE

AI Engineering Intern — VegaVisionary

Feb 2026 – Aug 2026

- Deployed a multi-agent RAG system (LangGraph) for AI-guided learning with mind maps and dynamic quizzes.
- Served 600+ students and 50 teachers; RAGAS scores of 80–88% across precision, recall, faithfulness, relevancy.
- Engineered PDF ingestion pipeline (FastAPI + Qdrant) with 3-tier hierarchical question mapping.
- Reduced 40-page PDF processing to under 1 minute in production, boosting teacher productivity by 70%.
- Built a real-time OMR evaluation pipeline, achieving 97% accuracy with latency of less than 3s.
- Built a hybrid image-generation pipeline to auto-generate diagrams and illustrations for AI-generated questions.

Analyst Intern — IIT Roorkee

May 2024 – Nov 2024

- Built Python–SQL pipelines processing 25K+ daily records, reducing manual effort by 80%.
- Designed LSTM and regression models achieving 87% accuracy for real-time forecasting.
- Integrated GenAI frameworks (LangChain, LangGraph, Gemini) for anomaly detection and automated reporting.

PROJECTS

MLOps Pipeline for Vehicle Insurance Cross-Sell Prediction



- Built end-to-end MLOps pipeline predicting customer interest in vehicle insurance using 12 features.
- Implemented automated CI/CD with GitHub Actions, Docker, AWS ECR, and EC2 for model deployment.
- Deployed FastAPI service with S3 model registry, MongoDB Atlas, and schema-driven data validation.

AgentX – Agentic AI Workspace — [Live](#)



- Built a production ReAct agent (LangGraph + FastAPI) with a sandboxed code executor and MCP tool support.
- Built a hybrid agentic RAG pipeline backed by a custom PDF parser for tables and images.
- Built AI cost-tracking and per-user credit limits, with LangSmith tracing on every LLM call.

TinyLLM – End-to-End Training and Alignment of Small Decoder-Only LM



- Built TinyLLM with WordPiece, RoPE, GQA attention, GELU, and pre-LayerNorm in a decoder-only model.
- Trained with mixed-precision next-token LM using AdamW with warmup for efficient single-GPU optimization.
- Applied instruction tuning and preference ranking alignment to refine model behaviour without RLHF.

Credit Card Score Modeling via Classification and Risk Techniques



- Built a predictive score using XGBoost on 30,000+ records, addressing class imbalance via SMOTE.
- Achieved F2-score of 0.603, identifying 84% of defaulters for early warning and loss mitigation.
- Engineered 8 financial features and tuned decision threshold aligned with the bank's risk policy.

Market Capitalization Forecasting with Deep Learning



- Built forecasting models to predict firm-level market capitalization across 1–3 year horizons.
- Developed LSTM, MLP, and Encoder–Decoder architectures trained on 25 years of historical market data.
- Applied PCA on 28 indicators to reduce dimensionality and improve model generalization.

POSITION OF RESPONSIBILITY

Web-D Joint Secretary — Team Wellness, IIT Roorkee

Jul 2024 – May 2025

- Managed and maintained the Wellness Centre website ensuring seamless user experience and timely content.
- Enhanced UI responsiveness, backend security, and accessibility through structured optimization.
- Conducted periodic audits and implemented data integrity measures for consistent reliability.