

Anubhav Verma

Data Scientist • Applied ML Engineer • NLP & GenAI Specialist
anubhavverma9369@gmail.com | +91-9451954022 | GitHub | LinkedIn | Tableau Public

PROFESSIONAL SUMMARY

M.Sc. Data Science candidate (VIT Chennai, CGPA: 8.56) with 4+ **production-deployed ML/NLP systems** spanning healthcare diagnostics, misinformation detection, and research grant analytics. Achieved **96% text classification accuracy** (BERT, 50k+ news articles) and **76.8% funding-decision accuracy** (XGBoost, 9,200+ grant records) — with every model shipped with explainability and a live deployment. DST INSPIRE Scholar (Top 1% nationwide). Seeking early-career **Data Scientist** roles to turn data into measurable business decisions.

TECHNICAL SKILLS

- **Programming & Databases:** Python, SQL, R, C, MongoDB, Hadoop
- **ML & Statistical Modeling:** Scikit-learn, XGBoost, TensorFlow, PyTorch, Regression, Classification, Clustering, Time-Series, A/B Testing, Hypothesis Testing, Cross-Validation
- **Data Analysis & Visualization:** Pandas, NumPy, Matplotlib, Seaborn, Power BI, Tableau, Excel, EDA, Feature Engineering, Statistical Inference
- **NLP & GenAI:** BERT, SciBERT, BART, Sentence-BERT, Hugging Face Transformers, RAG Pipelines, LLMs, FAISS, Pinecone, Fine-Tuning, Prompt Engineering, OCR
- **MLOps & Deployment:** Flask, Streamlit, Gradio, REST APIs, Git/GitHub, Vercel, Render, CI/CD Pipelines
- **Explainability & Analytics:** SHAP (XAI), Grad-CAM, Model Interpretability, Bias Detection, Vertex AI, Groq API

PROJECTS

Real-Time Misinformation Detection Intelligence Platform | *BERT, Flask, JavaScript, Deep Learning* Sep 2025

- Trained a **production-grade text classification model** on 50k+ multi-source news articles — **fine-tuning BERT** to achieve **96% accuracy** across politically and topically diverse datasets, with robust generalisation beyond training distribution.
- Reduced time-to-prediction from batch to real-time by **deploying the model via a Python/Flask REST API** consumed by a **JavaScript browser extension**, enabling **sub-second inference** at the point of reading with no user friction.

AI-Powered Research Grant Evaluation & Monitoring System | *SciBERT, XGBoost, SHAP, Streamlit* Nov 2025

- Delivered **76.8% funding-decision accuracy across 9,200+ grant records** by engineering a **hybrid NLP + gradient boosting pipeline** — SciBERT semantic embeddings fed into XGBoost — with **SHAP decomposition** making every funding prediction explainable and defensible to stakeholders.
- Converted model outputs into operational value by **deploying a real-time Streamlit dashboard** that tracks milestone progress, **flags fund misuse anomalies**, and visualises allocation metrics — giving program administrators a single pane of glass for portfolio oversight.

PaperMind: LLM-Powered Research Paper Q&A with RAG | *LLaMA 3.1, Groq, Pinecone, Sentence-Transformers, PyMuPDF* Mar 2026

- Solved the hallucination problem in document Q&A by building a **retrieval-augmented generation pipeline** — semantic chunking → 384-dim embeddings → **Pinecone vector search** → top-5 grounded context → **LLaMA 3.1** — reducing factual errors versus a prompt-only baseline.
- Validated the system end-to-end by **deploying a live Streamlit application** with chunk-score transparency, confirming retrieval quality in production and giving non-technical users a usable research assistant with zero setup.

EDUCATION

M.Sc. Data Science

Vellore Institute of Technology, Chennai, Tamil Nadu

Coursework: *Machine Learning, Deep Learning, NLP, Statistical Modelling, Big Data Analytics*

2024 – Present

CGPA: 8.56

B.Sc. Physics, Chemistry & Mathematics (PCM)

Dr. Rammanohar Lohia Avadh University, Ayodhya, UP

Coursework: *Linear Algebra, Probability & Statistics, Calculus, Classical Mechanics, Quantum Physics, Analytical Chemistry*

2021 – 2024

CGPA: 8.1

ACHIEVEMENTS & CERTIFICATIONS

- **DST INSPIRE Scholarship** — Government of India 2021 – Present
Full merit scholarship for Top 1% nationwide STEM performance; sustained 3+ consecutive years.
- **NPTEL Elite Certification: Introduction to LLMs** — IIT Madras Oct 2025
12-week MoE-funded advanced course on large language models; **Elite distinction (72%)**.
- **Generative AI Certification** — Intel Unnati Programme Aug 2025
Industry-sponsored GenAI certification: foundation models, prompt engineering, and deployment pipelines.
- **Basics of Quantum Information** — IBM Quantum Mar 2026
IBM-issued badge covering quantum states, circuits, and information-theoretic foundations.

SOFT SKILLS & LANGUAGES

Soft Skills: Strong Analytical Thinking, Clear Communication of Complex Ideas, Fast Learner, Comfortable Working with Ambiguous Problems, Collaborative Team Player, Attention to Detail

Languages: Hindi (Native), English (Professional)