RAHUL KUMAR SHARMA

BATHUA, GOPALGANJ, BIHAR, INDIA (841427) ; ✆ +919508476508

✉ kumarsharmarahul01@gmail.com 🔗 linkedin.com/in/rahul-kumar-sharma-aa0b57233 ⌂ github.com/Rahul29999

Education

Indian Institute of Technology (ISM) Dhanbad　　　　　　　　　　　　**Oct 2021 – May 2025**
Bachelor of Technology in Mining Engineering　　　　　　　　　　　　　*Dhanbad, Jharkhand*

- **Relevant Coursework:** Artificial Intelligence, Machine Learning (Supervised & Unsupervised), Deep Learning, Natural Language Processing, Data Visualization, Data Analysis, Probability & Statistics, Linear Algebra, Data Structures and Algorithms, Object-Oriented Programming, Python Programming language

Experience

Outlier AI (United States)　　　　　　　　　　　　　　　　　　**Nov 2024 – Jun 2025**
Data Scientist (LLM, Prompt Engineering) – Freelance　　　　　　　　　　　*Remote – India*

- Contributed to **RLHF-driven training and evaluation** of Large Language Models by ranking responses, refining prompts, and providing structured feedback to improve factual accuracy, coherence, and safety.
- Developed scalable **LLM evaluation workflows in Python** using OpenAI APIs and custom metrics, enhancing reliability of code generation and ML pipelines by over **15%**.

IBM　　　　　　　　　　　　　　　　　　　　　　　　　　**Jul 2025 – Aug 2025**
Open Source Contributor – Responsible Prompting API　　　　　　　　　　　　*Remote – India*

- Contributed to the **Responsible Prompting API**, enhancing embedding pipeline consistency and adding support for **Hugging Face models**, enabling safer, reliable **LLM interactions**.
- Improved **repository maintainability** by standardizing embedding filenames, resolving bugs, and refining documentation, boosting **cross-platform compatibility** and alignment with **responsible AI practices**.

EXTION　　　　　　　　　　　　　　　　　　　　　　　**Feb 2025 – May 2025**
Machine Learning Intern　　　　　　　　　　　　　　　　　　　*Uttar Pradesh, India*

- Developed an **NLP pipeline** in **Python** using **NLTK, VADER, Hugging Face Transformers, and PyTorch** to classify financial news headlines into positive, neutral, and negative sentiments with **92%+ accuracy**.
- Applied **text preprocessing** (tokenization, stopword removal, lemmatization) and created **visualizations** with Matplotlib/Seaborn to generate actionable insights into **market sentiment trends**.
- Optimized **GPU-accelerated inference** (CUDA/cuDNN) reducing model latency by **40%** and enabling real-time analysis of over **10K+ headlines**.

CodSoft　　　　　　　　　　　　　　　　　　　　　　　　**Aug 2025 – Sep 2025**
Software Development Intern　　　　　　　　　　　　　　　　　　*Remote – India*

- Built Java applications including **ATM Interface, Student Management System, and Currency Converter**, applying **OOP, file handling, and exception handling** concepts.
- Designed and implemented **modular code** for core operations like **transaction management, student record processing, and currency conversion** with proper input validation.

Projects & Research Work

Prediction of Longwall Periodic Weighing and Local Fall Intervals Using ML (Research Project) ⌂ [GitHub]

- Conducted a **research project** aimed at enhancing **mine safety and hazard prevention** by predicting roof fall intervals and periodic weighing patterns under given geo-mining conditions.
- Applied ML-driven analysis (**clustering, correlation, time-series, heatmaps, Mean Load Density trends**) on shield pressure data; identified **20+ periodic weighing events** (avg. span ≈ 22 m), a major **main fall of 54 m**, and multiple **local roof fall zones**, providing actionable insights for early-warning systems

Smart GenAI PDF Chatbot | RAG, FAISS, GPT-3.5, Sentence Transformers ⏻ GitHub
- Built a **PDF-aware chatbot** using **FAISS** and **Sentence Transformers** for semantic query retrieval.
- Integrated **GPT-3.5 fallback** to complete a **Retrieval-Augmented Generation (RAG)** pipeline for low-confidence answers.
- Enabled **real-time CLI Q&A** for educational, legal, and enterprise document automation use cases.

Bhagavad Gita NLP Chatbot | Semantic Search, Multilingual NLP, Gradio, FAISS ⏻ GitHub
- Built a **multilingual chatbot** using **Sentence-BERT** and **FAISS** for **semantic search** over 700+ Bhagavad Gita verses in Hindi, English, and Sanskrit
- Deployed with **Gradio** on **Hugging Face Spaces** to enable real-time interaction without any local setup or API key
- Applied advanced NLP techniques to deliver **90%+ accurate** philosophical Q&A; earned a **Kaggle Bronze Medal**

NTT Stock Price Prediction | ML, LSTM, Time-Series Forecasting ⏻ GitHub
- Built and evaluated a hybrid stock forecasting pipeline using 8+ classical regression models (**Linear Regression, CatBoost, Random Forest**, etc.) via **Scikit-learn**; achieved up to **95.46% R^2** score.
- Developed a deep **LSTM model with TensorFlow (Keras)** for time-series prediction, forecasting a **30-day price downtrend** ($178.97 $\rightarrow$ $102.04) with an **RMSE of 3.09**.

Technical Skills

Languages: Python, SQL
AI/ML: Supervised & Unsupervised Learning, Deep Learning, NLP, LLMs, Prompt Engineering, RAG, Semantic Search (FAISS, SBERT), Time-Series Forecasting
Frameworks/Libraries: Scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, FAISS, Sentence-Transformers, Pandas, NumPy, Matplotlib, Seaborn
Deployment & Tools: Git/GitHub, Docker, GitHub Actions (CI/CD), Streamlit, Gradio, Hugging Face Spaces, AWS (Basic), Azure App Service

Achievements

- **Kaggle:** 3× Bronze Medals – NLP Chatbot, E-Retail Segmentation, CNN Image Classification
- **LeetCode:** 1400+ rating, 200+ problems solved
- **Codeforces:** 1800 rating, 50+ problems solved
- **CodeChef:** 1700 rating, 120+ problems solved
- **HackerRank:** 4 in Python & Java, 400+ coding challenges solved

Leadership / Extracurricular

- Organizer – Khanan'22 Geo-Mining Fest, IIT Dhanbad
- Completed **BCG X Data Science Virtual Experience Program (Forage)**
- Volunteered with **Kartavya NGO** (teaching underprivileged children)
- Cricket & Badminton enthusiast