

# Sadika Sood

Senior Data Engineer | Bioinformatic Data Engineer | Machine Learning Expert | Clinical Genome Analyst

Web Scraping | Data Analysis | Full stack developer| LLM Engineer · \$30/hr · Available

Uttar Pradesh, India · [www.linkedin.com/in/sadika-verma](https://www.linkedin.com/in/sadika-verma)

## SKILLS

---

python, machine learning, deep learning, LLM, Agentic AI, SQL, NOSQL

## ABOUT

---

I am a Machine Learning and Bioinformatics professional with over 5 years of experience across industry and high-level research. Expert in developing computational pipelines for infectious disease control and genomic analysis. Recently specialized in Generative AI and Large Language Models (LLMs), contributing to model alignment through RLHF and Prompt Engineering via the Outlier (Scale AI) platform. Proven track record in Natural Language Processing (NLP), information retrieval, and full-stack development using Python, SQL, and AWS. Additionally, brings 1.5 years

## EXPERIENCE

---

### Freelance · Scale AI

Aug 2024 – May 2025

Remote · part time

Curated and validated large-scale training datasets for LLMs across Python and SQL reasoning tasks, improving model code-generation accuracy.

• Applied Reinforcement Learning from Human Feedback (RLHF) to refine model reasoning, reduce hallucinations, and improve instruction adherence.

• Evaluated model outputs using structured benchmarking frameworks measuring factual accuracy, logical reasoning, and computational

### Genome Analyst · 4basecare

Sep 2025 – Mar 2026

Remote · part time

Designed an automated clinical variant prioritization pipeline in Python to analyze cancer mutation datasets including SNVs (VCF), CNVs, and molecular biomarkers.

• Integrated multiple knowledge bases (ClinVar, VarSome, cBioPortal) to classify variants into ACMG clinical significance tiers.

• Implemented multi-stage filtering logic to accelerate identification of actionable genomic alterations.

### Bioinformatics Data Engineer I | TeOra · TeOra

Jun 2023 – Mar 2024

Bangalore · full time

Engineered a multi-stage RNAi candidate discovery pipeline, implementing consensus ranking algorithms and strict positional constraints for siRNA/shRNA design.

• Developed a biomedical semantic search system using ontology-driven query expansion and BioBERT embeddings, improving literature retrieval accuracy.

• Authored technical documentation and Standard Operating Procedures (SOPs) for end-to-end discovery workflows.

## EDUCATION

---

### University of Hyderabad · Diploma in AI and ML, Computer Science

Mar 2021 – Apr 2022

Hyderabad, IN

Data Analysis, Machine and Deep Learning

Java, Database etc