

Ruchi Sharma

+91 7073350932 - ruchi.0911.sharma@gmail.com - [linkedin.com/in/rsharma1109](https://www.linkedin.com/in/rsharma1109) - github.com/ruchi1109

SUMMARY

Experienced Data Scientist/ML Engineer with 3+ years building production-grade machine learning solutions, specializing in LLMs, GenAI, and data automation. Proven track record of delivering business-critical ML models that reduce processing time by 95 % while maintaining high quality standards. Strong focus on practical, iterative development with expertise in Python, R, cloud environments, and model deployment pipelines.

WORK EXPERIENCE

Associate Data Scientist

GlaxoSmithKline, Bangalore, India

May 2023 - Present

- **Independently developed** end-to-end **open-source R package** for automated pharmaceutical submission data validation, designed for **CRAN repository** publication. Built complete **Shiny application** performing quality checks using flexible pattern matching for clinical trial data formats and medical terminology standards. Created comprehensive **testing framework** and **cross-platform deployment** solution for **regulatory compliance** workflows.
- Developed an AI tool that used **GenAI** approaches like **GAN** and **VAEs**, focusing on synthetic data generation utilizing CT-GAN, ADS-GAN, and TVAE models, and conducted a comparative analysis against real data to evaluate model performance and accuracy.
- Developed a tool using the **XGBoost ML model**, powered by a **large language model**, which marked a pivotal shift from a fully manual approach to an automated one, reducing development time from two weeks to one hour.
- Developed an end-to-end system to identify user intents in short texts, using **unsupervised learning** leveraging techniques such as clustering sentence embeddings, **hyperparameter tuning of UMAP + HDBSCAN**, and automatic **cluster labeling**.

Junior Data Scientist

GlaxoSmithKline, Bangalore, India

Feb 2022 - April 2023

- Led a project to collate data from diverse sources and implemented **semantic search functionality** using the **Reinforcement Learning**, **Haystack framework** and **RoBERTa language model** resulting in a 50 percent reduction in man-hours.
- Developed a Python-based solution utilizing the **LayoutLM** model to detect Personally Identifiable Information (PII) in documents by implementing advanced preprocessing and tokenization techniques, enhancing automated detection and reinforcing data security compliance.
- Developed a **Power BI** dashboard integrated with a web-based data entry system, providing a single source of verified information, reducing validation time, and enabling faster data-driven decisions.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL

ML/AI Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn

LLM & GenAI: Prompt Engineering, RAG implementation, Haystack, Transformers (BERT, RoBERTa, LayoutLM), GANs, VAEs

Data Analysis & Manipulation: Pandas, NumPy, OpenCV

Data Engineering: SQL, Data Pipeline Development, ETL processes

Cloud & Deployment: AWS (EC2, S3), Docker, Git, Model Inferencing, ONNX familiarity

Workflow & MLOps: Agile methodology, Version Control, CI/CD basics, Production ML deployment

Visualization Communication: Power BI, Data Storytelling, Technical Writing

EDUCATION

Birla Institute of Technology, Mesra

B.E. (Electrical and Electronics)

Ranchi, Jharkhand, India

July 2014 - June 2018

EXTRACURRICULAR ACTIVITIES

- **Volunteer at Leaders for Tomorrow, India** - a not-for-profit youth leadership movement - June 2015 - Present
- **Cultural Committee member at BIT, Mesra** - Organized events, conducted workshops and Connected with Sponsors August 2014 - June 2018