

# RAJAN KUMAR SONI

9839828806

rajansoni424@gmail.com

leetcode.com/u/rajansoni424

linkedin.com/in/rajan-kumar-soni-unigalso

github.com/rajanskumarsoni

## WORK EXPERIENCE(4+ YEARS AFTER MASTERS)

### Machine Learning Engineer | @Phenom People Pvt.Ltd, Hyderabad

July 2023 – present

- Developed a recommendation engine using collaborative filtering and Graph Neural Networks, boosting conversion by 50% with 2M daily requests, utilizing MLflow, AWS EventBridge, S3, SageMaker, and AWS Lambda. Deployed for 400 clients, training the model daily using Kafka.
- Automated anomaly detection in CPU and memory usage, reducing alerts by 20% across 50 production services.
- Fine tuned and deployed a LLaMA-based ranking model for the recommendation system, outperforming the LTR model by 10% in Mean Reciprocal Rank metric.
- Developed and deployed a sequential recommender system (Bert4Rec) in production, improving apply-click rates by 14% over the Elasticsearch-based system.
- Built an AI agent using AutoGen to generate monthly business impact reports, reducing human effort by 50%.
- Deployed an AI agent using RAG and a vector database(chroma) for root cause analysis, reducing debugging time by 30%.
- Built sentiment and escalation prediction models for support tickets using (Large Language Model)LLM and Logistic Regression, reducing costs by 80%.

### Lead AI Engineer | @ImagoAI Pvt.Ltd,Gurgaun

Nov 2022 – July 2023

- Worked on Hyperspectral images (small dataset, skewed target distribution, high dimensional) to predict corn mycotoxins, corn damage, corn particle size, and corn starch present in them using xgboost, classical ML techniques, and variational autoencoder.
- Increased the revenue by 50%.
- Extracted the spectral-spatial features from HSI images using image segmentation.
- Managed a team of size 3 which includes assigning tasks and daily updates.
- Built the MLOps pipeline using various tools and technologies, such as AWS S3, databases, APIs, DVC, Datapane, and MLflow.

### Senior Data Scientist(III) | @Turtle shell technologies Pvt.Ltd(Dozee), Bangalore

May 2022 – Nov 2022

- Developed brain tumor detection model using image segmentation on MRI images with 95% accuracy.
- Built blood pressure prediction system using state-of-the-art ML models,using graph neural network, XGBoost to capture the relation between different vital signs in BCG signals and deployed the model on the server.
- Developed an early warning system where we generate alerts for different vitals using LSTM. Fine-tuned vitals threshold for alert firing using minimax optimization.

### Application developerI | @Oracle Financial Services Pvt. Ltd

july 2021 – May 2022

- "Created service extensions for the banking platform using Node.js, Java, and Embedded JavaScript templating.

## RESEARCH PUBLICATIONS

We propose task-agnostic embeddings for HMMs using a Graph Variational Autoencoder and class-label aware embeddings through diff pooling. These methods aim to effectively capture both the structure and behavior of HMMs.

- Soni, Rajan Kumar and Seshadri, Karthick and Ravindran, Balaraman** "Metric Learning for comparison of HMMs using Graph Neural Networks." *Proceedings of Machine Learning Research*, 2021, pp. 1365-1380.

<https://proceedings.mlr.press/v157/soni21a.html>

## EDUCATION

Program	Institution	CGPA	Year Of Completion
M.S. By Research (CSE) in ML and DS	Indian Institute of Technology, Madras	8.00/10	2022
B.Tech (Computer Science & Engineering)	MMMUT,UP	7.71/10	2018

## SKILLS

**Programming Languages:** Python, JavaScript, Java

**Machine Learning Frameworks:**PyTorch, TensorFlow, Scikit-learn, Huggingface

**Databases:** Snowflake, MongoDB, ElasticSearch, Vector Databases

**Cloud Platforms:** AWS (SageMaker, Lambda, S3)

**Tools:** MLflow, LangChain, AutoGen, Git, Docker