

MOHIT RAJPUT

ML | DL | AI | ML Engineer | Data Scientist | Data Engineer | IIT Roorkee | 8+ yrs Exp.
[LinkedIn](#) | [GitHub](#) | mohitrajput901@gmail.com | +91-8979572630

Full resume: <https://mr901.github.io/resume/>



Areas of Professional Interest and Experience

- Data Product Management, Data Solution Design, Solution Architecture, Data Collection, Data Engineering, R&D, Deployment, Monitoring and Maintenance (E2E)
- MLOps with Central and Decentralized Architecture with interest in EDGE processing
- Generative AI, Augmented LLMs, Agentic AI Systems, MCP Host – Servers Architecture, and Prompt Engineering
- Statistical, Machine and Deep Learning; Computer Vision, Natural Language Processing (NLP)

Work Experience

NerdAppLabs (engineering partner of Dianomic Systems)

Lead Machine Learning Engineer

Noida
Jun'23-Present

- Developed an LLM-powered autonomous reasoning MCP platform (MVP) integrating Excel, FogLAMPs, and Cloud – featuring cloud-side orchestration for multi-server coordination and MCP-driven Excel automation with direct access from FogLAMPs.
- Delivered FogLAMP MCP Server (v1) for centralized configuration, debugging, and analytics; Excel Datalink Add-on for real-time data connectivity; and built Medallion-architecture pipeline on Azure (Delta Lake + ADF) integrated with FogLAMP.
- Led end-to-end lifecycle of ML and Edge-AI solutions — from research, design, and development to deployment, monitoring, and maintenance across GCP, Azure, and on-prem environments.
- Architected and scaled in-house MLOps infrastructure, enabling cloud-agnostic and EDGE-ready workflows for data ingestion, training, deployment, versioning, and monitoring.
- Built intelligent signal-processing and predictive ML systems for centrifuge operations, including:
 - **Multi-Reference Anomaly Detector (POC)**: Designed Siamese and Contrastive Learning-based models on high-dimensional vibration data for operator-action and machine-response anomaly detection.
 - **Breakover Forecasting (EDGE Deployment)**: Developed and deployed Stacked-LSTM-based predictive model integrated with PLCs and HMIs via FogLAMP to minimize downtime and optimize production.
 - **Solids Build-Up Detection (EDGE Deployment)**: Implemented Autoencoder-driven time-frequency domain analysis for proactive maintenance.
- Engineered modular services for data collection, backfill, preprocessing, training, A/B testing, evaluation, and dashboard integration, with automated orchestration via Jenkins and Cloud Scheduler.
- Established a data warehouse and monitoring dashboards (Grafana, Postgres) to support ETL, OLAP, and unified visualization of data, model, and system health metrics.
- Authored comprehensive documentation using Sphinx and ReadTheDocs for internal developer enablement.
- **Tags: LLM, MCP, Model Context Protocol, FastAPI, FogLAMP/Fledge, Edge Compute, Predictive Maintenance, Anomaly Detection, Time-Series Forecasting, Signal Processing, Fourier Analysis, Autoencoder, Siamese Networks, Contrastive Learning, Multi-Instance Learning, LSTM, PLC, HMI, Cloud-Agnostic Architecture, MLOps, Model Monitoring, Data Versioning, CI/CD, Docker, GCP (Compute Engine, Pub/Sub, Cloud Functions, Cloud Storage, BigQuery), Azure (Data Factory, Delta Lake, Synapse), Grafana, Postgres, ETL, OLAP, Jenkins, Cloud Scheduler, AVEVA PI Server, PI Vision, Office.js, Excel Add-in, SSE, Documentation (Sphinx, ReadTheDocs).**

Candor Health

Senior Machine Learning Engineer

Bangalore
Mar'23-Jun'23

National Provider Identifier (NPI) Filler services for the missing NPI (incomplete-project)

- Filled in the missing information for multiple cases of Healthcare provider by utilizing the data available from data source.
- Tags: Identity Resolution, Similarity Mapping, Search, Parallel Processing, Big Data, APIs, Redis, Spark**

Amazon Seller Services Pvt Ltd - Analytics Technology and Engineering (ATE)

Data Scientist II
Engineer

Bangalore
Apr'22-Jan'23
Feb'23-Mar'23

Day Ahead Capacity Planning (Supply Chain)

- Providing volume forecast for last mile, to delivery stations to decrease under and over booking of logistics
- Tags: Time Series, ML, Sagemaker, EC2, S3, Monitoring**

Common Utilities Python Lib (V1)

- A small library for credentials accessing, SQL jobs creation and analysis, logger, s3 interaction, operations, clustering, data retrieval from quip documents, error calculations, data analysis and transformation utilities.
- Tags: AWS services, DB, SQL, Logging, Analysis, Sync Quip**

Capacity Planning Performance Evaluation & Descriptive Analysis

- Evaluate performance and provide multiple insights related to BAU, peaks, plans for current state and area for improvement, possible areas having data drift
- Tags: Time Series, Data Drift, EC2, S3, Excel**

Cargill - Cargill Digital Labs

Machine Learning Software Engineer

Bangalore
Jun'19-Apr'22

ML labs v2 setup

- Defining ways to Work with ML projects and other applications having ML related features
- Bringing sustainable practices for accelerated development while maintaining the agility
- Planning phase of data governance, 3rd Party Data Annotators, Portal development for data tagging, selection and versioning, data collection experiment designs and pipelines.

Estimating Growth Indicator of Chicken (POC then MVP)

- Estimating the weight and age of the chicken based on video recording done using the mobile camera.
Tags: Agile Experiment Setup & Design, Object Detection, Image Standardizer, Estimator, Ensembling

Defective Nugget Detection and Accounting (POC then MVP)

- Defective Nugget Detection and Tracking on conveyor belt with output projected on display for manual removal.
Tags: Jetson Nano, Raspberry Pi, Arduino Uno, Edge device, Object Detection, Tracking, on-premise, real-time

HR Chatbot (Myco) (MVP)

- Initial product version was launched for Singapore and Australia with capability for form filling, ticket creation, ticket checking, easy data integration from excel sheet
Tags: Rasa, Redis, NLU, NDM, NLG, NLP, Chatbot, MS Excel

Financial Commentary Generation (POC)

- A utility for table-to-text based on template method was developed for the initial stage, b/c of the lack and too much fragmentation of data source this approach was deemed fit.

YOLO v5-v4 training and deployment integration on AWS suitable to multiple tailored needs.

Tags: Computer Vision, Sagemaker, ECR, S3, Lambda Function, API Gateway, IAM, CloudWatch

Information Extraction from Ship Invoices (POC)

- Extracting multiple information through the information that was shared in old and current emails.
Tags: Text-to-Table, Analytics; Regex, Python Email to DB

Other

- **Streamlit** based Computer Vision Utility for parsing, labelling, visualizing annotation, with support for Yolo threshold adjustment for object detector and evaluate performance.
- MiApp feature involving weather data visualization using the **GIS data**.

ShieldSquare (Radware) - R&D

Data Scientist – R&D

Associate Data Scientist – R&D

Bangalore

Apr'18-May'19

Jun'17-Mar'18

Iterative Clustering and Learning based System to Separate traffic Apart (ICLSSTA)

- Solely developed a framework, which develop and evaluate unsupervised learning solution in much depth and this from a single config file.
Tags: Framework, Clustering, Anomaly, Outlier, Conceptual Drift Detection, Ensembling, Support all Dimension reduction, Anomaly, and Clustering Algorithm from Sklearn, Add. Custom Algorithm.

Adaptive Action Taking (AAT)

- Developed a generalized module named which makes use of reinforcement learning to automatically take action on incoming traffic on the web property.
Tags: Reinforcement Learning, Multi-Armed Bandit, Smart handling of Business Limits, Self-Adjusting

Deep Behaviour Analysis (IDBA)

- Partially developed a module named Intent based, which utilizes LSTM to yield encoded features and scores that are used with supervised, anomaly and clustering.
Tags: RNN, Auto-Encoder, Semi-Supervised Learning, Sequence Based Detection, Deep Learning

Threshold Online Reinforcement (ThOR)

- Helped developed a module named, which makes utilizes online learning to develop a probability score distribution over of isotonic regression for action taking.
Tags: Isotonic Regression, Online Learning, Self-Adjusting, Smart handling of Business Limits

Second Level Module Integration

- Developed and implemented end to end machine and reinforce learning based solutions using ICLSSTA-AAT, IDBA-ICLSSTA-AAT, & IDBA-ThOR
Tags: Semi-Supervised, Supervised, unsupervised, reinforcement learning, behaviour analysis.

Titan Batch Analyser

- Developed and Implemented, which utilizes multiple browser signatures, behaviour based rules, regular expressions and databases to generate a suspiciousness score to take action on traffic.
Tags: SQL & Python Implementation, ETL, Advance SQL operations, Handling Multiple Datasets

Rule Scripts

- Developed and Implemented multiple rules which were based on network, device and browser fingerprint, and behaviour of the visitors. These were adequately balanced between causal and correlated, yielding negligible FP.
Tags: Domain knowledge, Behaviour Analysis, Rules development

Rule Mining

- Created a Recommendation System for the highlighting possible bad signature coming in the traffic.
Tags: Associate Rule Mining, Text Data Pre-processing

Other

- Developed Dashboard, Visualization and Analysis, spreadsheet to understand the whole behaviour of the traffic. Partially interactive visualization sheets were also developed in python.
Tags: SQL and Spreadsheet/Python Implementation, Visualization, Result Sharing
- Developing Dynamic Moving **Collective Intelligence** which can evolve while working across multiple sid.

Python Package Developed

Algorithmic Machine Learning Exploration & Exploitation Tool (AMLEET) (personal)

- A big python package to accelerate the development, analysis, data wrangling, engineering and debugging process.
- It has support for **logging, configuration, notification, cloud storage integration, database connection, API creation, code functionalities, error capturing, performance logging, optimized file handling, parameter optimization etc**
- Other application involve generalized capability to data handling (standardization), analytics and science

- High level support for **computer vision** is also there with functionalities include video stream creation, interact with data lake, prediction wrapper for face detection, recognition, pose estimation and such; synthetic dataset creation.
- Additional supports for **NLP**, data cleaning pipeline, supervised, unsupervised ML, performance evaluation & optimization.

MCP-plots [[pypi link](#)][[smithery link](#)][[glamma link](#)][[mcp repository](#)]

- A Model Context Protocol (MCP) server for data visualization. It exposes tools to render charts (line, bar, pie, etc) from data and returns the plot as image / base64 text / mermaid diagram

COCO Transformation Utility (CTU) [[pypi link](#)]

- A python package to perform the same **transformation to coco-annotation** as performed on the image.
- Provides the capability to create augmented images and annotations.
- Capability of visualizing multiple kinds of annotations on images.

System and Network Information (SylInfo) [[pypi link](#)]

- A python package for gathering system information including hardware, specifications, network configuration, devices n network, real time system and process monitoring capabilities with support for querying of syslogs and installed packages.

Patent & Research Paper

- **First Author.** "System and method for detecting bots based on iterative clustering and feedback-driven adaptive learning techniques"; us US20200099713A1.
- **Second Author.** "System and method for detecting bots using semi-supervised deep learning techniques"; us US20200099714A1.
- **Second Author.** "Electric field and current assisted alignment of CNT inside polymer matrix and its effect on electrical and mechanical properties", in International Journal for the Science and Technology of Polymers 2016.

Education

Indian Institute of Technology, Roorkee

Jul'12-Jul'17

B.Tech. + M.Tech. (Integrated Dual Degree Course) in Metallurgy and Material Engineering, CGPA: 7.152/10
Ministry of Human Resource Development (MHRD) Scholarship for M.Tech. as for qualifying GATE.

Udacity Nano Degrees

- Data Product Manager, 2021
- Computer Vision, 2020
- Artificial Intelligence, 2018

Certified Courses from Coursera, Edx, Udemy

Details on this link <https://mr901.github.io/resume/>

Skills and Tools

Query Language: SQL (Advance)

Programming Language: Python (Advance), R (Intermediate) , Mojo (Beginner), JAVA (Beginner), C++ (Beginner) , SAS (Beginner)

Shell Scripting Language: Unix/Linux Shell, PowerShell

ML Packages/Framework: LangChain, MCP, scikit-learn, TensorFlow, PyTorch, numpy, pandas, streamlit, opencv, Dask, Flask
Other: FogLAMP, Fledge, Azure Data Factory, Google Big Query, Cloud Functions, Pub/Sub, Compute Engine, Redis, APIs, Docker, Microservices Architecture, Kubernetes (Beginner), Selenium, Jenkins, Apache Spark, Apache Kafka, PostgresDB , Grafana, HTML, CSS

Databases: SQL, No-SQL, Delta Lake

Software: Tableau, Excel, XLMiner, Jupyter Notebook, RStudio, Gimp, SolidWorks

Cloud Computing Services: Google Cloud Platform, Amazon Web Services, Microsoft Azure, Databricks, AVEVA PI System

Cloud Data lakes: GCS, S3, ADLS, Cloud Storage, Google-Drive, pcloud

Hardware: Jetson Nano, Raspberry Pi, Arduino Uno, IP Cameras, Sensors, Programmable Logic Controller (PLC), basic electrical devices and circuits

Languages: English(SRW), Hindi(SRW)

Projects

Mart chatbot available over whatsapp for searching, filtering and comparing the products, cart building and management and checkout and invoice functionality availability. Deployed on GCP with Kubernetes Event driven Autoscaling. (LLM powered) MCP server for plotting multiple types of charts and is now available across multiple providers.

Developed a Firefox-Extension that convert Medium article to a defined template to enable ease of reading and printing. [link](#)

Emails fetcher, filtering and summarizing based on LLM workflow developer for Gmail.

Getting video streams from IP camera to cloud. RTSP protocol to **HTML Live Stream** with Kafka streams for frames.

Accounting for Pipe Bundles being Exported [demo link](#)

Edge Device based CV application for counting of sheets processed in a workshop [demo link](#)

Sticker Detection and OCR [demo link](#)

Image to Image Search – Jewellery [demo link](#), Flicker8k [demo link](#)

Face Detection, Face Landmark Detection and Face Recognition Wrappers [demo link](#)

Human Pose Estimation and Activity Recognition [demo link](#)

WhatsApp based **Chatbot** (Hack around way)

Selenium based web scrapping and Web Application Control of Tinder & WhatsApp

Streamlit based threshold adjustment for object detector and annotation visualizer

Additional project details are present on this link <https://mr901.github.io/resume/>