MOHIT RAJPUT

ML | DL | AI | Data Collection | Data Engineering | ML Experiment Design | 6+yrs Exp.
LinkedIn | GitHub | mohitrajput901@gmail.com | +91-8979572630
Full resume: https://mr901.github.io/resume/

Areas of Professional Interest

- Data Product Management, Architect, Collection, Engineering, Research, Deployment, and Maintenance (E2E)
- MLOps with Central and Decentralized Architecture
- EDGE devices based solutions
- Statistical, Machine and Deep Learning; Natural Language Processing (NLP), ChatBots & Computer Vision

Work Experience

Candor Health
Senior Machine Learning Engineer — Mar'23-Present

National Provider Identifier (NPI) Filler services for the missing NPI
- 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, Redis, Spark

Amazon Seller Services Pvt Ltd - Analytics Technology and Engineering (ATE)
Data Scientist II — Apr'22-Jan'23
Engineer — 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 — Jun'19-Apr'22

ML labs v2 setup
- Defining ways to Work with ML projects and other applications having ML related features
- Brining 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, Ardino 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 — Apr'18-May'19
Associate Data Scientist – R&D — 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.

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.

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	Certified Courses from Coursera, Edx, Udemy
- Data Product Manager, 2021 - Computer Vision, 2020 - Artificial Intelligence, 2018	Details on this link https://mr901.github.io/resume/

Skills and Tools

Query Language: SQL (Intermediate)
Programming Language: Python (Advance), R (Intermediate), SAS (Beginner), JAVA (Beginner), C++ (Beginner)
Shell Scripting Language: Unix/Linux Shell, PowerShell
ML Packages/Framework: Scikit-learn, TensorFlow, PyTorch, Dask, Flask
Other: Google Big Query, Redis, APIs, Docker, Microservices Architecture, Kubernetes (Beginner), Selenium, Apache Spark, Apache Kafka, Grafana, Kibana, HTML, CSS
Databases: SQL, No-SQL
Software: Tableau, Excel, XLMiner, Jupyter Notebook, RStudio, Gimp, SolidWorks
Cloud Computing Services: Google Cloud Platform, Amazon Web Services
Cloud Data lakes: S3, Cloud Storage, Drive, pcloud
Hardware: Jetson Nano, Raspberry Pi, Arduino Uno, IP Cameras, Sensors, basic electrical devices and circuits
Languages: English(SRW), Hindi(SRW)

Projects

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/