Rohan Satya Isaac

rohan.isaac27@gmail.com • +1 765 631 5616 • Atlanta, GA • Portfolio • LinkedIn

SUMMARY

Data Scientist with experience in applying advanced analytics across a variety of sectors such as healthcare, manufacturing, CPG and Hi-Tech showcasing a strong ability to adapt and grow. Skilled in end-to-end data science from data curation, analysis, model building, evaluation and deployment

WORK EXPERIENCE

Emory University Hospital Atlanta, Georgia

Senior Data Analyst *Jul 2024 – Current*

- Built a data pipeline for a **large-scale multi-modality breast imaging dataset** of (2D and DBT mammograms, Ultrasound, MRI) for more than 1 million exams. Integrated multiple data sources such as patient workflow EHR data, risk scores, state registry data, clinical (encounters, labs, meds, notes, ICD, CPT), radiology and pathology reports extracted from EHRs. Tools: Python, Pandas, Polars, Epic, SQL, HL7, FHIR, Regex – Abstract presented in SIIM 2025
- **Evaluated a commercially deployed 3D DBT breast-AI screening model**, leveraging internal datasets to perform stratified analyses across outcomes, demographics, imaging characteristics and statistical evaluations to determine model optimization opportunities. Utilized Pandas for data transformations and Matplotlib, Seaborn, Plotly for visualization - paper
- Leverage medical imaging **foundation models** (MammoCLIP, MedImageInsights, BiomedParse) to extract screening mammography embeddings and applied linear probing for downstream breast imaging and cancer recurrence outcome analysis
- Built a **breast-segmentation model** for a tool that detects inadequate imaging in screening mammography at the point of acquisition, reducing need for technical recalls. Tools: PyTorch, Segformer
- Developed a module for **structured data extraction** from pathology reports by deploying **open-source reasoning LLMs** with a **pydantic-based schema** to extract over 30 labels spanning cancer outcomes, treatment details, and receptor status, achieving over 90% sensitivity. Applied **prompt engineering** techniques to improve parsing robustness and consistency. Tools: vllm
- Conducted multiple **clinical analytics research** for internal stakeholders specifically in breast imaging space, leveraging multiple data sources to generate data driven insights. Applied statistical methods to evaluate clinical research questions and inform evidence-based outcomes

ClearObject Remote

Data Scientist Intern *Jan 2024 – Apr 2024*

- Integrated computer vision systems in manufacturing quality control and monitoring, **involving data collection, labeling, and model fine-tuning using PyTorch and TensorFlow**, focusing on **segmentation (U-net), object detection (YOLO).** Tools: Google Vertex AI studio for end-to-end model development, finetuning and deployment
- Deployed models on **proprietary edge device** for real-time inferencing pipelined with **NVIDIA's TensorRT & DeepStream** to enhance operational frame rate efficiency.

Google Inc. – LibreHealth Radiology for Google Summer of Code Remote

Google Summer of Code Contributor – Work Product Report *May 2023 – Aug 2023*

- Integrated **computer vision** models to LibreHealth's **Radiology AI application** for medical imaging diagnosis tasks such as multilabel classification and object localization using GradCam based technique for chest and mammogram modalities.
- Automated the **AI model selection for an image's modality** by interacting it's DICOM properties cutting down the need to configure the specific AI model that will be used to inference on a specific modality example, CheXpert for Chest X-rays.

Indiana University – Health Informatics Lab Indianapolis, IN

Research Assistant *Sep 2022 – Apr 2024*

- Led an NSF study researching AI-Radiologist collaboration through which collected medical imaging data from multiple publicly available datasets, preprocessing them and management. Created inference scripts of **medical imaging computer vision** models for chest X-rays, Mammograms and Musco skeletal X-rays and **deployed** on MD.ai, a Radiology platform, reducing diagnostic times by 20%. Repo
- Contributed to the development of an **open-source radiology application's** AI system focused on integrating **medical imaging models for Chest and Breast modalities**. Merge Requests

- Built a **RAG** system using **LangChain**, Neo4j Graph DB, and ChatGPT API over clinical EHR data for semantic search and summarization.

Accenture Bengaluru, India

Data Scientist *Apr 2021 - Jun 2022*

- Improved a CPG client's **demand forecast** accuracy by ~15% using **Transformers and feature engineering**. Integrated into Azure Databricks Data Science workflow, and utilized **MLflow** to track experiments, log metrics, and compare experiment runs.
- Enhanced supply chain visibility through the design and development of manufacturing output metrics using **Power BI dashboards** and **Azure data pipelines** by bringing manufacturing data together. Leveraged Databricks (PySpark, SparkSQL), DataFactory, Synapse, and Tabular models for data processing and storage.

Senior Analytics Consultant *Apr 2020 - Mar 2021*

- Optimized **demand forecasting** accuracy for a Home Appliances client by 5% over base by integrating an **ensemble** of tree-based machine learning algorithms deployed to **Google Cloud**. Tailored **feature engineering** of promotions, pricing, and seasonality data.
- Developed, and **scheduled complex SQL queries** to run on Google BigQuery through Informatica ETL pipelines to extract and process data from data lakes and loads to data marts.

Analytics Consultant *Apr 2019 - Mar 2020*

- Developed and implemented a sourcing optimization solution, utilizing **text analytics and NLP** text similarity matching, boosting production efficiency by 20% for a key CPG client amid capacity challenges.
- For a Hi-Tech Client, developed, and integrated **large scale data preprocessing** scripts using sparkSQL, PySpark and Pandas on Azure that pulls sales order data and ML demand forecast data from data warehouse to build a supply forecasting solution.

Tata Consultancy Services Chennai, India

Systems Engineer *Oct 2016 - Jun 2018*

- Managed **Oracle ERP** technical service requests and developed **complex SQL queries** for custom reports for a High-Tech client.

EDUCATION

Master of Science, Applied Data Science
Indiana University-Indianapolis, (GPA: 3.9/4.0) *Aug 2022 - May 2024*

Post Graduate Program Data Science
Praxis Business School, Bengaluru, India *Jul 2018 – Mar 2019*

Bachelor of Engineering, Mechanical
Anna University-Chennai, India *Oct 2012 - Jun 2016*

SKILLS

Languages & Libraries: Python, R, Pandas, SciKit-Learn, TensorFlow, PyTorch, PySpark, Keras, NumPy
Generative AI Tools: LangChain, Vector/Embedding DBs (Pinecone, Chroma), Semantic Search, LLM Finetuning
Predictive Modeling: Boosting Regression/ Classification (LightGBM, XGBoost, Random Forest, LDA), Neural Network Architectures (CNN, U-Net, YOLO, Transformers, Foundational Models), Forecasting, Statistical ML – Linear, Logistic Regression
Engineering: ETL pipeline development, containerization (Docker / Kubernetes), Git, REST APIs, Distributed model training
Visualization: PowerBI, Matplotlib, Seaborn; **Databases**: SQL (MySQL, SQLite, BigQuery), NoSQL (MongoDB)
Cloud Tools: Google Cloud Platform (Cloud Run, Cloud Functions, BigQuery, VertexAI, Firebase), Azure (Databricks, Synapse, DataFactory)

PUBLICATIONS & OTHERS

- A general-purpose AI assistant embedded in an open-source radiology information system - 1st Author, AIME 2023 Conference
- Measuring Impact of Radiologist-AI Collaboration: Efficiency, Accuracy, and Clinical Impact – 1st Author, ISBI 2024 Conference
- Subgroup Performance of a Commercial Digital Breast Tomosynthesis Model for Breast Cancer Detection – 1st Author
- Kaggle Competition Runner Up – “SPR Screening Mammography Recall” Challenge
- Accenture – Google Cloud Hackathon (2022) – 2nd Runner Up