

RAM BADANA

Data Scientist

Email: ramhemanth580@gmail.com

Mobile: +91-6303774781

Linkedin: [@Ram Badana](#)

Huggingface: [Ram HF profile](#)

GitHub: [Ram580](#)

Summary

Highly motivated and results-oriented Data Analyst with 2+ years of experience driving data-driven decision making through advanced analytics, ML, Deep learning, NLP, and Generative AI solutions. Proven ability to design, develop, and deploy effective ETL pipelines, build and operationalize predictive models, and translate complex data insights into actionable strategies. Proficient in optimizing processes and leveraging advanced ML techniques to consistently deliver impactful results. Adept at various domains – (Pharma, Healthcare, Airlines). quickly adaptable to new technologies.

Skills

- **Programming Language & Packages:** Python (Tensorflow, keras, Pytorch, Transformers, Sklearn, NLTK), SQL, PySpark
- **Data Science:** Machine Learning (Regression, Classification, Clustering), Deep Learning (ANN, CNN, Sequence models), Natural Language Processing (NLP), Computer Vision (Image segmentation, Object detection), Time series forecasting, Data Visualization, Predictive Analysis
- **Data Analytics:** Statistics, Probability, EDA, Hypothesis Testing, ETL Pipeline development, Advanced Excel
- **Generative AI:** LLMs, Finetuning using PEFT Techniques (LORA, QLORA), RLHF, RAG applications, RAG Evaluation Techniques (RAGAS), Vector DBs, NoSQL DBs, Langchain – (chains, Tools, Agents), LlamaIndex, LLMOPS, Hugging Face
- **Technology & Cloud Platform:** AWS, Amazon Bedrock , AWS Sagemaker, Data bricks(beginner), Docker, Github Actions, FastAPI, Flask Rest API, Git
- **Big data analytics:** Spark, Hadoop
- **Data Visualization and Dashboarding:** Matplotlib, Seaborn , Plotly , Power BI , Tableau

Work EXPERIENCE

Data Analyst2 - TheMathCompany, Bengaluru, India

June 2022 – Current

- Developed a comprehensive Power BI dashboard suite leveraging Power BI, SQL, python, PySpark to analyse patient and HCP persistence, churn rates and measured marketing campaigns performance by creating over **300** KPIs, metrics. integrated data from 4 cross functional workstreams and **5** data warehouses with over **120 Mn** patients interactions from multiple brands enabling informed decision making and improved persistence rates.
- Leveraged SQL, Python, and PySpark to automate ETL Pipelines, which Efficiently streamlined data integration from the Datawarehouse reducing the turnaround time from 3-4 days to **4 hours** per month.
- Developed a cutting-edge Gen AI solution by creating a comprehensive knowledge repository for healthcare and pharma data by curating diverse sources (reports, Presentations, and data dictionaries) using Unstructured, Langchain (Chains, tools, Agents), Pinecone DB and OpenAI and developed Multimodal RAG system that transformed raw information into actionable insights, resulting in a **50%** reduction in data retrieval time and enabled informed decision making.
- Streamlined ETL pipeline that integrates enormous volume of data from Snowflake, CDL, Teradata data warehouses. And enhanced the performance of Prescriber Choice Model that determines the impact of multiple prescription touch-points for HCP's , by implementing ensemble techniques which boosted the model performance by **7%**.

- Led a team of 5 data analytics interns to build, operationalize and deploy multiple ML and DL models for various regression, classification, and clustering use cases with Accuracy ranging **88% to 94%**.
- Developed and Deployed Inventory Management app that forecasts passenger footfall using various time series forecasting models (ARIMAX, SARIMAX, FB Prophet, LSTM), with an **MAPE of 94%** for a prominent airline company.
- Trained an XGBoost model with **96% precision** for predicting flight price drops, leading to a substantial **10% increase** in booking conversion rates.
- Extracted raw data from Twitter APIs and analysed tweets to generate analysis showing trends in public opinion regarding policy changes. Delivered actionable trend insights, enhancing strategic decision-making and policy responsiveness by **25%**.

Projects

Natural Language to SQL generator and Data analyser

- Developed a Natural language to SQL generation and data extraction agent, powered with RAG Technique that empowers the prompt template with relevant examples and Table schema descriptions dynamically based on user input. Leveraged Langchain (chains, Agents, tools), LCEL, Chroma DB, Gemini Pro LLM, MySQL, AWS RDS, Huggingface spaces.

Realtime Face Mask Detector Using MobileNetV2

- Developed and Deployed Real time Image Classification web app that analyses images to detect facemask with 90% Accuracy. leveraging OpenCV, TensorFlow, MobileNetV2 and Streamlit.

Customer Service Agent using GPT 3.5 turbo

- Developed a customer service chatbot for ecommerce website leveraging the Open AI API , Langchain, Streamlit to automate workflows and enhance efficiency in handling user queries.

Youtube Video Summarizer Application

- Developed and Deployed a webapp in Hugging face spaces that can extract Video transcript and summarize to create a detailed notes using Finetuned LLAMA2 LLM.

Conversation Chatbot RAG Application

- Developed and deployed multiple pdf conversation agents, Multi Modal RAG Agents, Text to SQL engines using Gemini pro, Groq inference engine, Langchain Agents, Tools and Lance DB, Pinecone vectorDB.

Education

Manipal University

2018 – 2022

B. Tech Automotive Engineering (Minor specialization in EV and Hybrid vehicle Technology)

- CGPA : 8.2

Accomplishments

- 2 X Ambition and Hunger Award – TheMathCompany.
- 3 X Appreciation Awards from Multiple Project clients Mathco.
- IBM certified Data scientist.
- Secured 2nd place in Mathco Hackathon

Certifications

- IBM Data science professional Certification.
- Deep learning Specialization by Deeplearning.AI.
- AWS Generative AI Using LLMs Certification.
- Azure Power BI developer.