

ROHIT MACHERLA

✉ rohitmacherla125@gmail.com [in/rohitmacherla125](https://www.linkedin.com/in/rohitmacherla125) [RohitMacherla3](#) [Portfolio](#)

CORE COMPETENCIES

- Data Analytics (Python, SQL)
- Data Engineering (AWS, GCP)
- AI/ML (TensorFlow, PyTorch)
- Generative AI (LangChain)
- Computer Vision (OpenCV)
- MLOps (MLflow)

WORK EXPERIENCE

- Rutgers University, Graduate Research Assistant** **Oct 2022 – Dec 2023**
- Analyzed media data using **topic modeling** to uncover hidden narratives. Achieved **10x clustering speedup** for KMeans in FAISS, obtained 150+ clusters with BERTopic
 - Optimized data integrity by preprocessing using NLP techniques, resulting in a **30% data reduction**
 - Analyzed gene sequence and gene expression datasets, designed algorithms to run on GPU using CUDA to calculate gene-drug interactions, and **identified the top 10%** cases of interest based on **statistical analysis** and **bioinformatics**
 - Parallelized processing over 64 cores of a remote server, resulting in an **80% reduction** in execution time
- Omdena, Machine Learning Engineer** **May 2023 – Aug 2023**
- Engineered a **recommendation system** leveraging content-based, collaborative filtering and NLP techniques
 - Collaborated with Data Scientists and followed agile methodology to conduct exploratory data analysis, cleaning, and imputation to enhance **data quality to 98%**, and employed neural networks to report a **94% f1-score**
 - Rolled out an ensemble model to AWS utilizing **Streamlit** and FastAPI, enhancing the **click-through rates by 33%**
- Deloitte Consulting, Data Engineer** **Aug 2020 – Jun 2022**
- **Led** a team of 4, seamlessly integrating **Databricks** with Collibra Catalog using the JDBC Spark driver, and automating metadata ingestion for 260+ schemas using Python scripting and Tidal jobs to **reduce manual effort by 99%**
 - Engineered an **ETL** pipeline for data processing from **AWS S3** data lake to **Redshift** data warehouse. Transformed the data with **AWS Glue** using crawlers and orchestrated the pipeline using **Apache Airflow**
 - Developed **SQL queries** to extract data from MYSQL, Oracle, and PostgreSQL, optimized for efficiency, and exposed them as APIs
 - Facilitated and actively contributed to the successful execution of **5 production releases**, supporting and clearing challenges

PROJECTS

- Generative AI Applications** | [Git](#) | [Demo](#) **May 2024**
- Developed AI agents and chatbots with a Streamlit-based interface where users can engage in natural language dialogue to pose questions and gain insights about the input PDF file(s) or anything in general.
 - Leveraged OpenAI's embeddings to process the files into a **FAISS vector database**, created a **RAG** pipeline in **LangChain** with **ChatGPT-3.5**, and used prompting techniques to reduce token size
- Human Emotion Detection** | [Git](#) | [Demo](#) **Apr 2024**
- Implemented image classification using ResNet50 **CNN** architecture building it from scratch to achieve an f1-score of 78.6%
 - Conducted **Transfer Learning** on EfficientNetB4 to **improve** the score to **83.3%**. Fine-tuned the **Vision Transformer** model to obtain an accuracy and f1-score of **90%**
 - **Quantized** the model to **reduce the size by 92%** and deployed the model as a web app using **Streamlit**
- Customer Churn Prediction** | [Git](#) | [Demo](#) **Feb 2024**
- Performed survival analysis to identify the likelihood of churn and built a tree-based classification model to predict churn
 - Implemented **SMOTEEN** and **hyperparameter** tuning to solve the class imbalance problem, optimized the AUC score by 15% to **98.9%**, and deployed the **XGBoost** model as a user-friendly web application using Streamlit
- Text Summarization - NLP** | [Git](#) | [Demo](#) **Dec 2023**
- Performed text summarization on wikiHow dataset from Hugging Face using pre-trained **BART** and **T5 LLMs**
 - **Prompt Engineering** techniques were used on the input text to **improve** the BLEU score **by 10%**
 - Adopted simple and **LoRA fine-tuning** techniques to improve the **performance** on BLEU score **by 209%**

TECHNICAL SKILLS

- | | | | | | | |
|------------------|---------------|----------------|------------|--------------|--------------|------------------|
| • Data Mining | • A/B Testing | • PySpark | • Power BI | • MongoDB | • AWS Athena | • Docker |
| • Statistics | • Pandas | • Scikit-Learn | • NLP/NLTK | • Databricks | • Streamlit | • Apache Airflow |
| • Data Wrangling | • Numpy | • Matplotlib | • R | • BigQuery | • FastAPI | • SageMaker |

EDUCATION

- Rutgers University-New Brunswick, NJ** **May 2024**
Master of Science in Data Science GPA: 3.65/4
- National Institute of Technology Kurukshetra, India** **May 2020**
Bachelor of Technology GPA: 8.47/10