

ROHIT MACHERLA

✉ rohitmacherla125@gmail.com [in/rohitmacherla125](#) [RohitMacherla3](#)  Portfolio

CORE COMPETENCIES

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

WORK EXPERIENCE

Travelport, Data Scientist

Aug 2024 – Present

- **Integrated ChatGPT** with Webex as a **serverless** bot using AWS **Lambda**, API Gateway, and Redis for company-wide use, leveraging event-driven workflows and optimized **caching** to enhance scalability and response efficiency
- **Modernized** legacy data warehousing by migrating **COBOL**-based processes to a Python system on Linux, utilizing **Kubernetes**, Helm Charts, and **Airflow** for efficient orchestration, and improving scalability
- Assisted in estimating the performance of a **caching solution** using **PySpark** on AWS EMR clusters, achieving a **Hit Rate of 72%**
- Developed an ETL pipeline using PySpark to process hotel data on **AWS EMR** clusters, automating execution with **AWS Lambda** and **EventBridge** for daily scheduled runs

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 and 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**

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

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 | • Databricks | • AWS Lambda | • Docker |
| • Statistics | • Pandas | • Scikit-Learn | • NLP/NLTK | • BigQuery | • Streamlit | • Apache Airflow |
| • Data Wrangling | • Numpy | • Matplotlib | • MongoDB | • AWS EMR | • FastAPI | • SageMaker |

EDUCATION

Rutgers University-New Brunswick, NJ

Master of Science in Data Science

May 2024

GPA: 3.65/4

National Institute of Technology Kurukshetra, India

Bachelor of Technology

May 2020

GPA: 8.47/10