

ROHIT MACHERLA

📞 940-594-9613 ✉️ rohitmacherla125@gmail.com [in/rohitmacherla125](https://www.linkedin.com/in/rohitmacherla125) [RohitMacherla3](https://github.com/RohitMacherla3) 📁 Portfolio

CORE COMPETENCIES

- Data Analytics (Python, SQL)
- Data Engineering (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** with KMeans in FAISS, uncovered 150+ clusters with BERTopic
- Optimized data integrity preprocessing using NLP techniques resulting in a **reduction of the data by 30%**
- Analyzed gene sequence and gene expression datasets, implemented algorithms to calculate gene-drug interactions, and **identified the top 10%** cases of interest based on **statistical analysis** and **bioinformatics**
- Enforced parallel 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** that leveraged content-based, collaborative filtering and NLP techniques
- Conducted data cleaning and statistical data imputation techniques to **enhance data quality to 98%**
- Explored matrix factorization and neural networks, to achieve a **94% f1-score**
- Deployed an ensemble model to AWS utilizing Streamlit and FastAPI to enhance the **click-through rates by 33%**

Deloitte Consulting, Data Engineer Aug 2020 – Jun 2022

- **Led** a team of 4, seamlessly integrated **Databricks** with Collibra Catalog using JDBC simba spark driver. Automated metadata ingestion for 260+ schemas using Python scripting and Tidal jobs to **reduce manual effort by 99%**
- Developed **SQL queries** to extract data from MYSQL, Oracle, and PostgreSQL, optimized for efficiency, and exposed them as APIs
- Engineered an **ETL** pipeline for data processing automation of Qlik Sense data into Informatica Cloud (IICS) and Collibra via REST API calls using Unix Script, parallelized the ingestion for a **66% time reduction**
- Facilitated and actively contributed to the successful execution of **5 production releases**

PROJECTS

PDF Chatbot with LangChain and OpenAI | [Git](#) | [Demo](#) May 2024

- Developed a chatbot with a Streamlit-based dynamic interface where users can engage in natural language dialogue to pose questions and gain insights about the input PDF file(s)
- Leveraged OpenAI's embeddings to process the files into a **FAISS vector store**, implemented 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%
- Performed **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 over time and built a tree-based classification model
- Implemented **SMOTEEN** and **hyperparameter** tuning to optimize 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
- Statistics
- A/B Testing
- Pandas
- Numpy
- PySpark
- Scikit-Learn
- NLP/NLTK
- R
- MongoDB
- Databricks
- BigQuery
- Streamlit
- FastAPI
- SageMaker
- AWS
- Airflow
- Kafka
- Matplotlib
- Power BI
- Informatica

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