ROHIT MACHERLA

📞 940-594-9613 ✉ rohitmacherla125@gmail.com 🔗 in/rohitmacherla125 ⯃ RohitMacherla3 💼 Portfolio

Education

Rutgers University-New Brunswick, NJ   **Graduating: May 2024**
Master of Science in Data Science   *GPA: 3.65/4*
Coursework: Regression & Time Series Analysis, Data Structures & Algorithms, Probability & Statistics, Data Mining, Statistical Modeling & Computing, Database Management, Statistical Software (NLP), Statistical Learning (DL)

National Institute of Technology Kurukshetra, India   **Aug 2016 – May 2020**
Bachelor of Technology   *GPA: 8.47/10*

Work Experience

Rutgers University, Graduate Research Assistant | Git   **Oct 2022 – Dec 2023**
- Optimized data integrity through data standardization across 3 sources and preprocessing using NLP techniques resulting in a **reduction** of the data **by 30%** and **removal** of URLs, HTML tags, and emojis **by 99%**
- Achieved **10x clustering speedup** with FAISS in KMeans, delving into DBSCAN, DP-Means, and ultimately opting for BERTopic, uncovering 150+ clusters
- Analyzed health datasets, designed and implemented Python algorithms to calculate gene-drug interactions, and **identified the top 10%** cases of interest based on estimated **statistical** parameters
- Enforced parallel 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 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 databases, optimizing efficiency, and seamlessly exposed the results as APIs using MuleSoft proxy for enhanced accessibility and integration
- 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**
- Designed **Power BI** Dashboards through Collibra APIs, showcasing asset metrics and metadata completeness, driving a **30% accuracy awareness boost**. Incorporated **3-layer drill-through** for enriched asset lineage comprehension
- Facilitated and actively contributed to the successful execution of **5 production releases**, demonstrating expertise in deploying and maintaining data engineering solutions to ensure operational stability

Omdena, Machine Learning Engineer | Git   **May 2023 – Aug 2023**
- Collected and curated crowdsourced data from over 75+ contributors, conducted EDA, and employed data cleaning and statistical imputation techniques to **enhance data quality** by achieving a **98% completion rate**
- Engineered a **recommendation system** that leveraged content-based, collaborative filtering and NLP techniques. Explored matrix factorization and neural networks, to achieve a **94% f1-score**
- Implemented an ensemble model, to enhance the **click-through rates by 33%**
- Deployed the models to AWS utilizing Streamlit and FastAPI for users to interact and test as a POC

Projects

Human Emotion Detection - Computer Vision | *Git*   **Jan 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%**. **Deployed** the model to **Heroku** using FastAPI endpoints

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%**

Twitter Search Application | Git   **Apr 2023**
- Extracted and transformed user and tweet data from JSON files, succeeded with **100% accuracy** in storage into MYSQL and MongoDB, leveraging analytics tools for EDA and sentiment analysis
- Developed search functionalities with custom **caching** techniques and attained a **500-fold enhancement**

Technical Skills

- **Proficient**: Python, SQL, Pandas, Numpy, Sklearn, TensorFlow, PyTorch, Matplotlib, GCP(Certified Data Engineer), Databricks, Collibra, ETL, Statistics
- **Worked with**: R, Unix, MongoDB, LLMs, BigQuery, Power BI, Informatica Cloud(IICS), Surprise, FastAPI
- **Basics**: Spark, SageMaker, MLflow, Tableau, spaCy, Streamlit, A/B testing, AWS