GAAJARLA SHESHUGOUD

Data Scientist | GenAi Engineer

📞 +91-8125391505 ✉ gaajarlasheshugoud@gmail.com in Linkedin ○ Github

PROFILE

Passionate Data Scientist with a deep understanding of machine learning and deep learning algorithms, and a passion for exploring the frontiers of generative AI through extensive research paper analysis. Specialized in Transformer architectures, my expertise lies in applying cutting-edge AI techniques to solve complex problems, driving data-informed strategies that propel organizational success.

EDUCATION

Osmania University   **2023**
*Post Graduation Diploma in Data Science (part-time) - **CGPA - 8.78***   *Hyderabad*

Jawaharlal Nehru University   **2018**
*Mphil, International Trade - **CGPA - 7.9***   *New Delhi*

TECHNICAL SKILLS

- Programming Langauges (Python,R)
- Machine Learning
- Deep Learning
- NLP
- GenAI

- Mlops
- LLmops
- Database (SQL, NoSQL and Vectordatabase)
- Tensorflow, Pytorch,

- Scikit-learn
- NLP(Transformer — BERT — GPT — ELMO — Roberta — Distilbert)
- Cloud

- Technologies(AWS, Azure,GCP)
- Tech Stack (Statistics, ETL, Transformers, LargeLanguage Models)

PROFESSIONAL EXPERIENCE

Tritech Software Pvt. Ltd. | Data Scientist   **March 2021 - Present**

- Developed a deep learning model for efficient and accurate language-to-language translation, maintaining original text's meaning and context.
- Curated a diverse dataset of parallel texts from various sources, encompassing multiple genres and styles for model training.
- Implemented an attention mechanism from scratch, using BLEU score as the primary performance metric to evaluate translation accuracy.

Perfexion Information Technologies Pvt. Ltd. | Data Analyst   **June 2018 - Feb 2021**

- Managed and optimized daily reports for 500+ delivery partners, enhancing report accuracy by 20% and maintaining a 95% on-time submission rate.
- Boosted query efficiency by 30%, refining over 50 SQL queries on Hubble and Data Central, improving response times significantly.
- Developed Excel-based sales reports, achieving 99.5% accuracy, which increased reporting process efficiency by 25%.
- Utilized Amazon QuickSight for dynamic analytics, leading to a 30% increase in sales productivity and a 40% improvement in data visualization.
- Played a key role in over 10 data-driven initiatives, enhancing delivery partner efficiency by 25% and customer satisfaction by 20%.

PROJECTS

Sensor Fault Detection [PG Diploma Data Science Project] ⬀

- Fetch Streaming data through Kafka and store it to MongoDB.
- Trained a Machine Learning Classification Models to predict accurate prediction and Deployed a web app on AWS EC2 with cloud watch monitoring.
- Designed a continuous training pipeline using GitHub actions on remote machines and utilized an S3 bucket as a model registry.
- Deployed Dockerized application using GitHub actions for faster inferencing.
- Tech Used: Python, Mongodb, Machine Learning, CICD, Github Actions, Docker, Fast API

Automatic Ticket Classification System with Integrated Question Answering ⬀

- ∗ Developed an Automatic Ticket Classification system leveraging advanced Natural Language Processing (NLP) techniques and machine learning algorithms to streamline support operations within the organization.
- ∗ Implemented a robust data pipeline for extracting and storing information from PDF documents using Pinecone vector database, ensuring efficient data retrieval and processing.
- ∗ Designed and trained a classification model using CSV data, employing data preprocessing techniques and machine learning algorithms to accurately categorize user queries and requests.
- ∗ Integrated a question answering component into the system, enabling users to receive immediate responses to their queries based on available data, thus enhancing user experience and reducing response times.

RAG-Based Chat Application for PDF Document Management ⬀

- ∗ Integrated a RAG framework enabling efficient categorization of PDF documents based on priority, enhancing document management capabilities.
- ∗ Developed functionality to interact seamlessly with multiple PDF files, allowing data extraction,Augmentation, and Generation within the application.
- ∗ Established connectivity with a vector database, facilitating quick access and retrieval of relevant information associated with PDF documents.
- ∗ Leveraged the FASSI framework to ensure scalability, performance, and responsiveness of the application, enhancing user experience.
- ∗ Tech Stack: Database-FASSI,, LangChain, OPENAI, RAG, Data (70 Pages PDF File), Streamlit

AutoGenMCQ: Automated Multiple-Choice Question Generation Tool with OPENAI ⬀

- ∗ The project aims to automate the process of generating MCQs from textual data, enhancing educational content creation and assessment practices. Leveraging the capabilities of OpenAl and Streamlit, the system provides users with a user-friendly interface to upload text data, generate MCQs, and download the output in CSV format.
- ∗ Addressing potential scalability and performance issues, especially when processing large volumes of text data or generating numerous MCQs simultaneously.
- ∗ Optimizing the user interface for seamless interaction and providing customization options to cater to users' preferences and requirements.

CERTIFICATIONS

- Data Science Boot Camp - Ineuron
- Stanford University - Machine Learning
- GenAi Boot Camp - Success Analytics
- Deeplearning.ai - AI for Everyone, Generative AI for Everyone, Genetrative AI with LLM's

EXTRACURRICULAR ACTIVITIES

- ∗ Columnist contributing to Telugu and English newspapers on a variety of contemporary issues, including technology.
- ∗ Reading novels and writing poetry are my priorities during leisure time.