Anusha Yaramala

USA, Texas-76207 | 9409771496 | anushayaramala31@gmail.com

PROFILE

Certified GenAI specialist and highly skilled Data Scientist with **4+ years** of extensive experience in designing and deploying MLOps pipelines and data processing using Bigdata and python. Proficient in leveraging LLM's, AWS and Azure and GCP services.

SKILLS

- **Programming Languages:** C, Java, R, C++, Linux, Python (Pandas, NumPy, Matplotlib, Seaborn, SciPy, OpenCV, MLFlow)
- **GenAI stack:** Multi-model RAG's, GPT-4, Claude, Co-pilot, Fast API, MCP, Semantic search, Vertex AI, Azure factory.
- **Cloud Tech Stack:** AWS (Glue, RDS, S3, Athena, EC2, RDS, Redshift, Lambda), Databricks, GCP, Azure DevOps, BigQuery, Airflow.
- **Big Data Ecosystem:** PySpark, Snowflake, Kafka, Flink, Data Governance, Data Modeling, Hadoop, Delta Lake, Delta Tables
- **Databases/Storage:** OLAP, OLTP, DBMS, PostgreSQL, MySQL, NoSQL, SQL Server, Mongo DB, Data Governance, Metadata Management
- **Machine Learning Algorithm:** ANN, Linear Regression, CNN, Logistic Regression, RNN, Decision Trees, SVM, Random Forests, Naive Bayes

EXPERIENCE

Bright house financial | GenAI Specialist **Sept 2024-present | USA**

- Working on RAG and BERT models with LLMs to enhance contextual understanding and extract the required aspect-based summaries, from Client policy documents using TensorFlow, Hugging Face transformers, and OpenAI to build chatbots reducing the manual calls by 40%.
- Worked with Agentic systems, vector databases like Pinecone and FAISS, fine-tuned Llama using RLHF, QLoRA for memory-efficient training, integrated with LangChain for dynamic prompt orchestration, achieving 70% reduction in GPU memory usage and 30% reduction in training time.
- Implemented NLP libraries such as NLTK, spaCy, scikit-learn optimizing text analysis, chunking and sentiment analysis. Utilized feedback mechanisms to train ML models, delivering more personalized user experiences.
- Presented complex data findings and insights from A/B tests and predictive analyses to clients, senior management, and stakeholders.

Texas Instruments | Neural network engineer **Feb 2023 - Jul 2024 | USA**

- Developed and optimized neural network architectures using SciKit Learn, Pandas, keras, pyTorch, numpy, matplotlib within Azure Databricks, achieving a 91.68% F1 score and reduced computational complexity by 35%.
- Used NLP algorithms, including Named Entity Recognition (NER) and Sentiment Analysis using VADER, improving text analysis efficiency by 70% and facilitating insight extraction from social media data.
- Designed, deployed scalable Deep learning models into production using APIs and pipelines improving deployment efficiency by 30%.
- Utilized Latent Dirichlet Allocation (LDA) for topic modeling, uncovering key themes and trends from textual data, with stemming and tokenization preprocessing techniques, and produced interactive visualizations with Plotly.
- Employed Azure DevOps and Git for version control and CI/CD best practices, while leveraging AWS SageMaker to automate workflows and manage infrastructure accelerating model scalability, training and deployment by 20%.

Deutsche bank | Data Engineer **May 2021 – Jul 2022 | India**

- Developed and optimized complex SQL queries and DDL scripts, improving query performance by 35% through index optimization, and executed queries for complex datasets exceeding 1 million rows.
- Optimized distributed data processing workflows using Python, PySpark, and Spark SQL, reducing data transformation time by 40% and enhancing predictive analytics and system efficiency by 20%.
- Implemented real-time data streaming with Google Cloud Pub/Sub and Apache Kafka, with low-latency processing of 10 TB+ of data.
- Utilized Hadoop Distributed File System (HDFS) and Spark SQL for efficient distributed storage and querying.
- Managed ETL pipelines handling 100K to 2M records with diverse schemas, achieving data cleaning and transformation in 2 to 10 min.
- Applied statistical modeling techniques such ROC, confusion matrices, chi-square and PCA for model evaluation and optimization.
- Optimized Airflow jobs to automate 100+ data workflows in GCP, enhancing operational efficiency by 25%.

Vodafone | Data Insights Analyst **Sept 2020 - May 2021 | India**

- Improved query running time by 40% by utilizing stored procedures in SQL instead of multiple joins. Managed procedures, installed databases, and solved problems in both development and production environments.
- Collaborated with cross-functional teams to develop data-driven dashboards enabling real-time performance tracking and decision-making.

PROJECTS

- **YouTube Trend Analysis:** Developed a scalable data pipeline on AWS to analyze YouTube video trends, processing 1 million+ records daily using S3, IAM, Quick Sight, Glue, Lambda, and Athena, Quick sight for data ingestion, transformation, storage, and visualization with 80% accuracy.
- **Publications:** 'Artificial neural network modelling of aluminium/Al2O3/fly ash hybrid composites prepared by powder metallurgy

EDUCATION

Master's in data science - University of north Texas 3.9GPA
Bachelor of Technology - Lakireddy Balireddy College of Engineering 75%

CERTIFICATIONS

Oracle cloud infrastructure GEN AI (Oracle) | AWS Certified Cloud Practitioner | Databricks Certified Developer