Teja Swaroop Sayya

Software Engineer | +1(980) 230 4200 | teja.sayya108@gmail.com | linkedin.com/in/teja-sayya/ | leetcode.com/Teja_Sayya/

EDUCATION

Masters of Science, Computer Science **May 2025**
University of North Carolina at Charlotte | GPA: 4.0/4.0

Bachelors of Technology, Computer Science **June 2021**
Jawaharlal Nehru Technological University| CGPA: 9/10

EXPERIENCE

Open Source Developer – GitHub & Google Developer Group **June 2023 - Present**

- Worked on backend improvements for the RocketChat, FitCheck App, optimizing DB queries and streamlining API integrations for scalability.
- Contributed to the TensorFlow Addons repository, implementing a custom loss function and fixing compatibility issues with TensorFlow

Software Engineer, Amazon(Client)-*Hyderabad* **Oct 2022 – May 2023**

- Engineered backend services using Spring Boot and implemented microservices architecture to ensure scalability and modularity for enterprise applications.
- Contributed to the development of an ML-based recommendation system, focusing on fine-tuning model parameters, optimizing data pipelines, and ensuring seamless data ingestion for accurate predictions.
- Assisted in integrating the recommendation system with RESTful APIs, enabling real-time personalized suggestions for users.
- Built advanced dashboards using ReactJS, visualizing key metrics and insights from the ML recommendation engine.
- Collaborated with data science and engineering teams to enhance system performance by refining data preprocessing steps and pipeline efficiency.

Big Data Engineer, CVS Health(Client)-*Hyderabad* **June 2021 – Oct 2022**

- Automated ETL workflows with Tableau Prep, reducing manual effort by 95%, saving 35+ hours monthly.
- Optimized 50+ Sqoop jobs, managing 5TB of data in HDFS and Hive, ensuring low-latency queries.
- Reduced query times by 60% on 10TB datasets using Spark RDD transformations and actions.
- Built data quality framework using Apache Spark for schema validation & data profiling on 1TB+ datasets, ensuring data accuracy & integrity.
- Applied advanced Spark transformations to manipulate and analyze unstructured data, executing complex text mining algorithms on 1 billion+ records for user table utilization insights.
- Managed end-to-end data pipelines in GCP, integrating PySpark and BigQuery to streamline data flows and mitigate risks in cloud-based infra.
- Extensively worked with Hadoop, PySpark, AWS EMR, and the broader Hadoop ecosystem (Hive, HDFS, Spark) to architect and manage big data solutions, ensuring scalability and performance optimization.

SKILLS

Programming:	**Data Structures and Algorithms (DSA)**, Java, Python, JavaScript, Golang
Web Frameworks:	Spring Boot, FastAPI, Vue.js, REST APIs, Microservices
AI/ML Libraries/Frameworks:	LangChain, LlamaIndex, RAG(Retrieval-Augmented Generation), Hugging Face, TensorFlow/Keras, PyTorch
Databases:	MySQL, MongoDB, BigQuery, ChromaDB
Big Data Frameworks:	Apache Spark, AWS EMR, Hadoop, HDFS, PySpark, MapReduce
Cloud Platforms:	Amazon Web Services (AWS), Google Cloud Platform (GCP)
Developer Tools:	Google Colab, Anaconda, CUDA, Docker

ACADEMIC PROJECTS

- **Chat with Teja Sayya | RAG(Retrieval Augmented Generation):** link
  Built a RAG-based AI agent using LLM functionality and custom data, enabling scalable and context-aware conversational AI Agent.
- **Connect4 with AI | Alpha-Beta Pruning Search Algorithm:** GitHub Repository
  Developed an AI-powered Connect Four game using Python and Pygame with an interactive GUI. Implemented various algorithms like Minimax and Alpha-Beta Pruning to create challenging AI opponents. Skills: Python programming, AI algorithms, and Search Algorithms via games.
- **Alzheimer's Disease Analysis | Machine Learning:** Github Repository
  Researched ML models to analyze the impact of various factors on Alzheimer's risk using Linear Regression, Random Forest, SVMs, and XGBoost.
- **MakeItTalk | Deep Learning models:** Hugging face spaces
  On research paper "MakeItTalk". Converts an Image & Audio file into Facial Animation Video. Used Convolution Neural Networks(CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), LSTM (Long Short-Term Memory) networks
- **Photo App | ReactJs:** link
  Developed a scalable React-based photo sharing application with MongoDB as the database and Material UI for the frontend.
- **Real-Time Chat Application | Spring Boot, Web Sockets + React**: GitHub Repository
  Developed a real-time chat application using Spring Boot, WebSockets, and React, enabling users to join chat rooms via unique room IDs and communicate seamlessly. Integrated MongoDB for data persistence and deployed containerized microservices using Docker.