Lalitha Samyuktha Jayanthi

+91 7569606146 samyukthajayanthi09@gmail.com Visakhapatnam

Value Proposition

Data Science graduate with a strong foundation in data science, big data processing, and statistical analysis. Adept at solving complex problems through scalable ML solutions and collaborative engineering. Demonstrated ability to lead cross-functional projects and communicate insights effectively. Seeking to contribute to data-driven decision-making in dynamic environments.

Education

- B. Tech | CSE-Data Science | Avanthi Institute of Engineering and Technology | 2025 | 7.88 CGPA
- Intermediate | Sri Chaitanya Jr. College | 2021 | 9.12 CGPA
- Tenth | Sri Chaitanya School | 2019 | 9.5 CGPA

Technical Proficiencies

- **Programming:** Python, SQL, Hive, Spark, TensorFlow, PyTorch
- **Big Data & Cloud:** AWS (EC2, S3, Lambda), Docker, Jenkins
- **Data Science:** Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn
- **Statistical Techniques:** Regression, t-test, chi-square test, hypothesis testing
- **Tools:** Git, GitHub Actions, Linux, Prometheus

Projects

1.Customer Churn Prediction Using Big Data
Tech Stack: PySpark, Hive, SQL, Scikit-learn, AWS S3 **Overview:**

- Ingested millions of customer records from Hive tables and S3 buckets
- Used PySpark for feature engineering and parallel processing
- Built a logistic regression model to predict churn probability
- Visualized churn trends using Matplotlib and Seaborn **Impact:** Demonstrates scalable data processing and predictive modeling.

2. Real-Time Stock Market Sentiment Analyzer
Tech Stack: Spark Streaming, Kafka, Python, SQL, Hugging Face Transformers **Overview:**

- Streamed tweets and news headlines using Kafka
- Processed data in real-time with Spark Streaming
- Applied NLP sentiment analysis using BERT
- Stored results in SQL database for dashboarding **Impact:** Combines big data pipelines with data science and NLP.

3. Healthcare Data Lake for Patient Analytics
Tech Stack: Hive, Spark SQL, Python, AWS Glue **Overview:**

- Created a data lake architecture for hospital records
- Used Hive for schema-on-read queries and Spark SQL for aggregations
- Built Python scripts to analyze patient readmission rates
- Applied statistical tests (t-test, chi-square) for insights **Impact:** Highlights data engineering, analytics, and statistical rigor.

4. Supply Chain Optimization with Distributed Data
Tech Stack: PySpark, SQL, Pandas, Docker **Overview:**

- Processed shipment logs from multiple regions using PySpark

- Built a clustering model to identify bottlenecks
- Used SQL for querying and Pandas for post-processing
- Containerized the pipeline with Docker for portability **Impact:** Shows ability to handle large datasets and optimize operations.

5. Big Data-Powered Resume Matcher
Tech Stack: Hive, Spark MLlib, Python, SQL **Overview:**

- Stored resumes and job descriptions in Hive tables
- Used Spark MLlib for TF-IDF vectorization and cosine similarity
- Built a Python interface to rank resumes based on job fit
- Logged results in SQL for audit and feedback **Impact:** Combines big data, ML, and real-world application

Internships & Simulations

- AI Intern – AIMER Society | 8 Weeks
- ML Intern – Indian Servers | 8 Weeks
- Data Science Intern – Yhills | 8 Weeks
- Data Science Intern – Prodigy | 4 Weeks
- Job Simulation – TCS | Forage Plat form

Certifications

- Google Cloud Gen AI.
- Data Analytics by Jobaaj Learnings.
- SQL by Oracle.
- Java by Oracle.