Adithya Singupati

(930) 333-2369, aditya.singupati@gmail.com, https://www.linkedin.com/in/adithyasingupati
https://github.com/MrAdithya21, https://mradithya21.github.io/mywebs.github.io/

Education

Indiana University Bloomington — *Master of Science in Data Science*
Graduation: May 2025
Relevant Coursework: Big Data Systems, Applied Machine Learning, Applied Algorithms, Data Engineering, Advanced Database Technologies, Data Visualization, Deep Learning, Natural Language Processing
Gayatri Vidya Parishad College of Engineering — *Bachelor of Engineering in Computer Science*
Graduation: June 2023 — GPA: 3.7/4.0
Relevant Coursework: Software Engineering, Big Data Analytics, Data Mining, Artificial Intelligence

Skills

Big Data Tools: Apache Spark, Databricks, AWS (S3, Lambda, SageMaker), Redis
Programming: Python, SQL, Scala, Java, JavaScript, R, C
Analytics & Data Engineering: Data Pipeline Development, ETL, Data Quality Checks, Dimensionality Reduction, Feature Engineering
Machine Learning: Classification, Clustering, Regression, NLP, Recommendation Systems, Topic Modeling, Deep Learning, PyTorch, TensorFlow, Model Deployment
Visualization & Reporting: Tableau, Power BI, D3.js, Matplotlib, Confluence Documentation
Collaboration & Tools: Git, JIRA, Agile, Cross-functional Team Communication

Projects

Graph-Based Identity Ingestion Pipeline Simulation
- Simulated a graph data ingestion pipeline that mimics production environments using Databricks and Spark.
- Automated quality checks, generated reports for ingestion verification, and documented results in Confluence.
- Implemented automated metrics validation scripts using PySpark and SQL to assess merge effectiveness.

Segmentation Analysis for Customer Insights
- Built Spark-based data pipelines to preprocess 1M+ customer records on Databricks and applied k-means clustering.
- Reduced dimensionality with PCA; integrated insights into Power BI to inform marketing strategies.
- Resulted in a 20% uplift in targeting accuracy for campaigns.

Real-Time Generative AI Chatbot with AWS Integration
- Developed a chatbot using Hugging Face transformers integrated with AWS Lambda and SageMaker.
- Deployed real-time inference endpoints and optimized monitoring pipelines.
- Improved response accuracy by 30% using feedback loops and retraining workflows.

Database System for Retail Inventory
- Designed scalable databases and built ETL pipelines to handle retail transaction data across multiple sources.
- Created Tableau dashboards with advanced filters and benchmarks to support business decision-making.
- Ensured data accuracy and consistency using automated checks.

News Article Topic Modeling and Classification
- Trained unsupervised LDA models on millions of news articles to extract latent topics.
- Built classification pipelines to categorize articles using TF-IDF, Word2Vec, and transformer embeddings.
- Tools: PyTorch, scikit-learn, spaCy, BERT, evaluated using F1-score and AUC.

Experience

Data Analyst — *O'Neill School of Public and Environmental Affairs*, IN, USA   Nov 2024 – Present
- Created real-time dashboards using D3.js and Databricks, presenting insights to over 1,000 stakeholders.
- Developed Spark-based pipelines for automating ingestion and quality control of large public datasets.
- Authored data summary and reporting documents on Confluence for auditing and collaborative reviews.

Data Scientist (Research Assistant) — *Kelley School of Business*, IN, USA   Mar 2024 – Present
- Processed 80M+ news articles using NLP to build sentiment scoring models supporting financial risk analysis.
- Created scalable pipelines on AWS and automated metric reporting for model validation.
- Deployed and monitored Spark-based transformations in AWS EMR and S3.

Data Scientist Intern — *MyEdMaster LLC*, VA, USA   May 2024 – Aug 2024
- Built scalable regression pipelines in Databricks using PySpark for SAT performance prediction.
- Automated data preparation workflows, reducing processing time by 40%.
- Worked closely with frontend developers to embed prediction APIs into live dashboards.

Publications

Tracking the Storm: Visualizing Trends in U.S. Hurricanes, IRJET, Vol. 12 Issue 1
Used big data and geospatial analytics to create hurricane impact dashboards.
IRJET-V12I138.pdf
Accessing General Health Care Facilities, IRJET, Vol. 9 Issue 10
Built a cloud-enabled e-health platform for real-time hospital inventory management.
IRJET-V9I10150.pdf

Certifications

- **AWS Certified Cloud Practitioner**
- **Tableau Desktop Specialist**