Satheesh Meadi

📞 +1 (667)-433-1382 | smeadi1@umbc.edu | Leetcode | LinkedIn | GitHub

EXPERIENCE

University of Maryland Baltimore county(UMBC)

Data Engineering Mentor   **May 2024 - Apr. 2025**

- Empowered mentees to deliver robust ETL pipelines as measured by a 40% increase in on-time project completions, by providing hands-on guidance in Python, Airflow, and PostgreSQL best practices.
- Elevated student proficiency in cloud data architectures by 50% as measured by pre- and post-training assessments, by conducting workshops on AWS (S3, Lambda, Redshift) and instructing on scalable data modeling techniques.

AI Researcher - Generative Models   **Sep. 2023 - May. 2024**

- Improved generation speed by 35% as measured by throughput tests, by fine-tuning transformer models (e.g., GPT-2, BART) in PyTorch and optimizing training routines on available GPU resources.
- Applied data augmentation and adjusted attention weights to cut text inconsistencies by 25% (measured via BLEU/ROUGE), resulting in noticeably more coherent outputs.

LTI Mind Tree

Data Engineer   **Jan 2023 - Aug. 2023**

- Boosted ETL throughput by 30% as measured by average job runtime reductions, by architecting and optimizing Spark-based ETL pipelines on AWS and implementing partitioning/indexing strategies in Snowflake and PostgreSQL.
- Accelerated data warehouse query performance by 5× as measured by benchmark times on critical dashboards, by migration SSAS cubes to a Snowflake data lake on GCP, redesigning star-schema models, and automating pipeline orchestration with Airflow.

Data Engineer Intern   **Sep. 2022 - Jan. 2023**

- Increased data ingestion throughput by 25% as measured by throughput logs, by developing Python and SQL-based ETL scripts, integrating diverse sources CSV/JSON into a Postgres data mart, and orchestrating workflows with Airflow.
- Reduced data quality errors by 30% as measured by anomaly-detection reports, by implementing validation checks and profiling routines in SQL and Python for incoming datasets, ensuring compliance with data governance standards.

Cognizant | Data Engineer Intern   **Aug. 2021 - Sep. 2022**

- Designed and deployed scalable AWS-based data pipelines (Lambda, EMR, Redshift, S3) to process high-volume insurance datasets, improving data update efficiency and system reliability by 60% through automation and robust data modeling.
- Developed machine learning models for insurance premium prediction, achieving 98% accuracy by engineering features from structured and unstructured policyholder data.

SKILLS

Languages: Python, Java, C++, SQL, Bash
Databases & Tools: MySQL, PostgreSQL, Snowflake, Git, GitHub, GitHub Actions, Docker, Kubernetes, Airflow, Streamlit
Big Data Frameworks: Apache Spark, PySpark, Hadoop, Kafka
Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly
Cloud: GCP, AWS (EC2, S3, Lambda, Glue, Redshift), Microsoft Azure(Data Factory, Synapse, Blob storage)
Data Science & ML: LLM's , CAG, RAG, Gen AI, Natural Language Processing(Spacy, NLTK), Hugging Face, MLflow, SageMaker

PROJECTS

Medi-Text Summarizer: Medical Text Analysis Application 🔗 Link

- Engineered a Streamlit-based medical text analysis system, boosting processing efficiency by 60%, by integrating PubMedBERT, NER, and DistilBERT to summarize 1000+ pages, extract 150+ entities, and analyze sentiment with 95% accuracy.
- Implemented multilingual support for 15+ languages via Google Cloud API, enhancing accessibility for 200+ healthcare professionals, by integrating real-time sentiment visualizations and interactive Q&A using Ollama's Llama3.2 model.

Financial Data Pipeline and Risk Analysis Dashboard 🔗 Link

- Automated ETL pipelines to ingest and process real-time financial data from APIs, improving data ingestion speed by 40%, by optimizing storage and retrieval in PostgreSQL and Snowflake for portfolio analysis and risk assessment.
- Developed a Streamlit dashboard for CAPM analysis, efficient frontiers, and risk-return models, enhancing financial data visualization efficiency by 35%, by integrating distributed data processing with Apache Spark.

Dribble, Predict, Win: Leveraging PySpark & ML for Real-Time NBA Game Forecasts: 🔗 Link

- Developed a large-scale PySpark data pipeline, enhancing real-time NBA game prediction efficiency by 50%, by processing 10,000+ player and team statistics.
- Optimized ML models (logistic regression, random forest), increasing prediction accuracy to 92%, by fine-tuning hyperparameters and feature engineering for sports analytics and forecasting.

EDUCATION

University Of Maryland Baltimore County   Maryland, MD

Masters of Professional Studies in Data Science; GPA: 3.967   May 2025