

Madhumitha Sekamuri

madhutrish06@gmail.com | +1-614-483-7876 | [LinkedIn](#) | [GitHub](#) | Ashburn, VA(open to relocation)

Summary

Senior Data Analyst with 3+ years of experience transforming complex datasets into actionable insights through SQL, Python, Tableau, and Power BI. Skilled in exploratory and scenario analysis, A/B testing, and predictive forecasting to improve decision-making and operational efficiency. Proven track record of building scalable dashboards, ensuring data quality, and delivering insights that align with business strategy and drive measurable outcomes.

Experience

Moxie
Senior Data Analyst

Ashburn, VA
Feb 2025 – Present

- Built and maintained Tableau and Power BI dashboards using SQL, BigQuery, and Python to solve visibility gaps in financial and operational KPIs, boosting reporting efficiency by 40% and enabling faster leadership decisions.
- Performed exploratory and scenario analysis on structured and unstructured data, applying statistical techniques to uncover trends and outliers, reducing reporting discrepancies by 25% and optimizing workflows.
- Developed predictive forecasting models in SQL and Excel for outbound volume planning, improving forecast accuracy by 15% and supporting more cost-efficient resource allocation.
- Led a stakeholder-driven data quality initiative, co-developing data profiling and monitoring tools with Airflow and Python; this reduced data error rates by 30%, increased trust in analytics outputs, and aligned reporting with SLAs and accessibility standards.

Behavioral Neuroscience – The Ohio State University

Columbus, OH
May 2024 – Dec 2024

Data Analyst Intern

- Analyzed high-frame-rate behavioral datasets by implementing a SLEAP-based tracking system in Python and TensorFlow on Google Cloud, improving annotation efficiency by 70% and enabling more accurate downstream analysis.
- Processed and structured sequential interaction data using CNN- and LSTM-based pipelines in PyTorch on GCP, generating labeled features that supported statistical analyses and improved classification accuracy to 92%.
- Leveraged transfer learning with U-Net models in Keras and TensorFlow on GCP TPUs to extract fine-grained behavioral signals, converting outputs into measurable engagement metrics with an F1-score of 0.92 on custom datasets.
- Conducted comparative analyses between baseline and fine-tuned models to evaluate incremental improvements, providing data-driven evidence of deep architecture benefits for experimental design decisions.
- Developed an automated analytics workflow in Python, NumPy, and Cloud Composer to streamline preprocessing, inference, and reporting, ensuring reproducible results and faster delivery of behavioral insights.

Pelotonia Research Center – Cancer Research

Columbus, OH
Jan 2023 – May 2024

Data Analyst

- Built and automated ETL pipelines in Python, SQL, and Apache Airflow on Azure Data Factory to ingest and clean large-scale EHR and genomic datasets, improving data accessibility and boosting reporting efficiency by 50%.
- Processed and standardized medical imaging datasets using MONAI and PyTorch on Azure ML, applying registration techniques (diffusion, affine, deformable) that increased alignment accuracy to 97%, enabling consistent downstream analysis.
- Conducted statistical modeling and survival analysis with scikit-learn, XGBoost, and Lifelines on Azure ML Studio, applying Kaplan-Meier and log-rank tests to uncover patient survival trends and treatment response patterns.

Developed interactive Tableau dashboards deployed via Azure App Service to visualize survival curves, treatment efficacy, and patient stratification, empowering clinicians and researchers with self-service insights for evidence-based decisions.

Cognizant Technology Services

Chennai, India
Mar 2021- Dec 2022

Machine Learning Engineer

- Built and deployed **risk prediction models** using **LightGBM and pandas** on **Amazon SageMaker** to classify high-risk insurance policies, enhancing fraud detection and underwriting decisions with a **30% improvement in accuracy** across 100K+ policy records.
- Deployed ML models as **REST APIs using FastAPI**, containerized with **Docker**, and hosted on **AWS ECS**, while managing experiment tracking and version control via **MLflow on S3**, streamlining the MLOps lifecycle in a regulated insurance environment.
- Designed and orchestrated **scalable data pipelines** using **Databricks (PySpark)** and **dbt**, integrated with **AWS Glue and Redshift**, automating ingestion and transformation of **1M+ insurance records** and reducing model retraining time by **40%**, supporting real-time policy risk scoring workflows.
- Conducted comparative benchmarking of machine learning algorithms (**LightGBM, XGBoost, Logistic Regression**) using cross-validation and **AUC/F1** metrics to determine optimal model for policy risk classification.
- Built and automated end-to-end **ETL pipelines** using **Python, SQL, Apache Airflow, and Pandas** on **Azure Data Factory**, to ingest, clean, and integrate large-scale structured and unstructured **EHR and genomic datasets**, improving data accessibility and reporting efficiency by **50%**.

Skills

Languages: Python, R, Java, JavaScript, C, D3.js, HTML/CSS
ML: Scikit-learn, XGBoost, LightGBM, Lifelines, H2O, PyTorch, TensorFlow, Keras, MONAI, CNNs, LSTMs, U-Net, LBPH, Transfer Learning, PyTorch Geometric
LLM & NLP: OpenAI (GPT-4, ChatGPT APIs), Google PaLM, LLaMA 2, Mistral, LangChain, Retrieval-Augmented Generation (RAG), Prompt Engineering, ChromaDB, FAISS, Pydantic, TextBlob
Data Analytics & Visualization: Tableau, Power BI, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Exploratory Data Analysis (EDA), Scenario Analysis, Cohort Analysis
Statistical Modeling & Forecasting: Regression, Time Series Forecasting, A/B Testing, Survival Analysis (Kaplan-Meier, Log-rank Test), Classification, Clustering, Predictive Analytics
Data Engineering & Pipelines: Apache Airflow, Azure Data Factory, BigQuery, AWS Glue, ETL Development, Data Cleaning, Data Integration, Data Quality Management
Databases & Cloud: Azure SQL, MySQL, PostgreSQL, Redshift, DynamoDB, AWS (Lambda, S3), GCP (AI Platform, BigQuery)
Machine Learning (for Analytics use cases): Scikit-learn, XGBoost, LightGBM, Lifelines
Data Privacy & Compliance: Data Governance (HIPAA, PII), Data Profiling, Data Quality Monitoring, Dashboard Development, Business Case Analysis
Certifications : Microsoft Certified – Azure Data Fundamentals, Google Data Analytics Professional Certificate, Tableau Desktop Specialist, AWS Certified Data Analytics – Specialty

Projects

AI-Powered Business Idea Evaluation System:
Built an end-to-end pipeline to evaluate 500+ business ideas using LLMs (LLaMA 2, Mistral, ChatGPT APIs) and classical ML models; performed text preprocessing, EDA, and model benchmarking with Logistic Regression and Decision Trees, achieving over 85% agreement with expert-labeled criteria through cross-validation and hyperparameter tuning.

Text-to-SQL App:
Built a **Streamlit application** that translates natural language into SQL queries using **LangChain and Google PaLM**, enabling seamless interaction with **MySQL databases**. Integrated prompt chaining, query execution, and result visualization to support real-time data exploration for business users.

Smart Mobility - [Smart-Mobility-GNN](#)
Built a GNN-based model using **PyTorch Geometric** on the **Open Traffic Dataset**, modeling road networks as graphs for route optimization and traffic prediction; achieved a **15% improvement in path prediction accuracy** over baseline models.

Face-X - [Face Recognition](#)
Developed a high-accuracy facial recognition attendance system using OpenCV, Haar Cascade, and LBPH, achieving 98% accuracy in real-time identification enhanced with Dlib and FaceNet for robust feature extraction. **Presented and Published at ICIVC 2022.**

Education

Master of Science in Computer Science and Engineering – Ohio State University – CGPA: 3.53

Dec 2024

Coursework: Neural Networks, Data Mining, Fairness in Artificial Intelligence and Databases, Data Visualization, Parallel Computing, Advanced OS, Cybersecurity

Bachelor of Technology in Computer Science and Engineering – Anna University – CGPA: 4

June 2021

Coursework: Problem-Solving, Object-Oriented Design, Data Structures & Algorithms, Databases, Operating Systems, System Design, Networking, Machine Learning