Sai Sindhu Nallapaneni

(972) 643 8285 saisindhunallapaneni1106@gmail.com Linkedin

Summary

Results-driven Data Analyst with 3+ years of experience in healthcare analytics, specializing in large-scale data analysis, AI model benchmarking, and value-based care optimization. Proven impact at CitiusTech and UnitedHealth Group through predictive modeling and advanced ETL development using Python, Spark, dbt, and AWS Glue. Expert in visualizing insights via Power BI and R Shiny, while ensuring HIPAA/FDA compliance and integrating standards like FHIR and HL7.

Technical Skills

Languages:	Python, SQL, R
Tools	Pandas, NumPy, Scikit-learn, Plotly Dash, SciPy, Statsmodels, GitHub, Confluence, Notion, Great Expectations, dbt, AWS Glue, AWS Lambda
Databases & ETL:	Apache Spark, Apache Airflow, PySpark, Apache Kafka, AWS Glue, dbt, Snowflake, Databricks, SQL, Neo4j, GraphQL
Data Visualization:	Power BI, R Shiny, Plotly Dash, Notion, Tableau
Methodologies:	A/B Testing, Feature Engineering, Dimensionality Reduction, Dimensional Modeling, Cross-Functional Collaboration, Data Storytelling, Regulatory Compliance (HIPAA, FDA)
Big Data Technologies:	Apache Spark, Apache Kafka, PySpark
Data Governance:	Great Expectations, Collibra, AWS Lake Formation, FHIR, HL7, LOINC, HEDIS, CMS Star Ratings

Experience

Data Analyst | United Health Group, USA **Feb 2025 – Present**

- Devised real-time analytics pipelines using Apache Kafka and AWS Lambda to monitor patient engagement with chronic care programs, accelerating care coordinator response times and improving adherence to treatment plans.
- Optimized data models in Snowflake and dbt to support cost-benefit analysis of Medicare Advantage enrollees across various care pathways, enabling finance teams to prioritize funding for high-impact interventions.
- Designed stratification logic using Python, SQL, and XGBoost to segment dual-eligible populations based on utilization patterns, which informed targeted care plans and drove a 12% increase in patient satisfaction scores.
- Built HIPAA-compliant patient journey maps using Neo4j and GraphQL to identify service gaps in preventive care, resulting in a redesign of outreach strategies that reduced care delays in underserved regions.
- Integrated social determinants of health (SDOH) datasets with claims and EHR data using AWS Glue and PySpark, uncovering key factors impacting medication adherence, which informed the rollout of a tailored health coaching initiative.
- Benchmarked provider performance using R Shiny dashboards linked to CMS star ratings and HEDIS scores, enabling provider networks to visualize care quality gaps and align incentives accordingly.
- Conducted A/B testing on member communication strategies using Python (SciPy, Statsmodels) to evaluate outreach effectiveness, while enforcing enterprise-level data governance policies via Collibra and AWS Lake Formation, ensuring both statistical rigor and regulatory compliance across campaign data.
- Facilitated cross-functional data storytelling sessions using Power BI and Notion, translating technical insights into business actions for clinical leads and executives, which guided quarterly quality improvement initiatives.

Data Analyst | CitiusTech, India **May 2020 – Apr 2023**

- Analyzed diverse healthcare datasets using Python (Pandas, NumPy) and Apache Spark to uncover data quality issues and standardization gaps across AI models, facilitating a 20% increase in Medictiv model interoperability.
- Engineered ETL pipelines using SQL, dbt, and Apache Airflow to aggregate model metadata from federated clinical systems, ensuring accurate and consistent data ingestion to support the open AI model directory.
- Visualized AI model performance metrics and patient cohort segmentation using Power BI and Plotly Dash, enabling data scientists and clinicians to compare efficacy across models and make evidence-based decisions.
- Synthesized clinical, claims, and genomic data using Snowflake and Databricks, driving integration of multimodal datasets to enhance the transparency and explain ability of models listed on Medictiv.
- Validated data pipelines and model evaluation outputs by designing automated data profiling scripts in Python and Great Expectations, reducing manual QA efforts by a 35% and boosting model reliability in production.
- Collaborated with various teams to map healthcare ontologies using FHIR, HL7, and LOINC standards, improving semantic consistency and enriching metadata quality across 100+ AI model entries.
- Curated datasets for benchmarking AI models by applying feature engineering and dimensionality reduction techniques using Scikit-learn, leading to the successful deployment of 15 new models in Medictiv's directory.
- Documented reproducible data workflows and metadata lineage using Confluence and GitHub, enhancing auditability and facilitating regulatory alignment for AI model governance in compliance with HIPAA and FDA guidelines.

Education & Certifications

Master's in computer science, Auburn University, AL, USA **May 2023 - Dec 2024**

Bachelor of Engineering in ECE, Panimalar Engineering College, Anna University, India **May 2017 - Apr 2021**

- Python Programming Foundation– Apponix Technologies
- Data Science with Python Workshop– Genesys Academy
- Python Programming for Everybody – University of Michigan, Coursera

Projects

Academic Research Project & Published Paper – IJSART Journal

- Published research in IJSART on designing a CNN-based brain tumor classifier using segmented MRI data, combining traditional image preprocessing in MATLAB with deep learning techniques to improve diagnostic accuracy.