VAMSI KRISHNA SAMAVEDAM

DATA ENGINEERING AND AI APPLICATION DEVELOPMENT

Manor, TX | (512) 757-0969 | samavedamvamsikrishna1@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Data-focused software engineer and M.S. Computer Science candidate experienced in Salesforce development at Accenture, entity-resolution research, and AI application development. Builds ETL/ELT pipelines, dimensional models, and data quality frameworks using Python, SQL, PySpark, Airflow, dbt, Snowflake, AWS (S3, Glue, Athena), and Azure (Data Factory, Synapse). Develops LLM-powered analytics applications (FastAPI, LangChain, natural language to SQL) with human-approved transformations and guarded SQL execution; applies machine learning to record linkage and anomaly detection.

TECHNICAL SKILLS

Languages and Databases: Python, SQL, PostgreSQL, SQL Server, Snowflake, DuckDB, Java, C++
Data and Cloud Engineering: ETL/ELT, PySpark, Apache Spark, AWS S3, AWS Glue, Amazon Athena, Azure Data Factory, ADLS Gen2, Databricks, Azure Synapse Analytics, Microsoft Fabric, OneLake, Delta Lake, dbt Core
Architecture and Delivery: Data Warehousing, Dimensional Modeling, Star Schema, Medallion Architecture, Incremental Loading, Data Quality, Data Validation, Apache Airflow, Docker, Git, Pipeline Monitoring
AI and Application Development: FastAPI, REST APIs, Pydantic, LangChain, LangGraph, Gemini API, LLM Workflows, Prompt Engineering, Natural Language to SQL, Pytest
Machine Learning and Analytics: Pandas, NumPy, scikit-learn, Splink, Entity Resolution, Logistic Regression, Random Forest, Neural Networks, Autoencoders, Anomaly Detection, Power BI
Salesforce: Apex, Triggers, SOQL, SOSL, Flow, Validation Rules, Lightning Components, Data Loader, Profiles, Permission Sets, Role Hierarchy, Organization-Wide Defaults

PROFESSIONAL EXPERIENCE

Graduate Research Assistant | Texas State University *San Marcos, TX | Jun 2026–Present*

- Developed an end-to-end probabilistic record-linkage and entity-resolution pipeline for 47,000+ historical persona records using Python, Pandas, Splink, and DuckDB.
- Engineered cleaning, standardization, blocking, and candidate-generation workflows across names, dates, locations, gender, and family relationships.
- Designed matching features using Jaro-Winkler similarity, temporal constraints, gender compatibility, and parent/spouse relationship evidence.
- Built and evaluated Logistic Regression and Random Forest classifiers using manually reviewed record pairs for duplicate-person prediction.
- Generated consolidated entity clusters and relationship datasets across baptism, marriage, and burial records for family-tree reconstruction and graph analysis.

Advanced Associate Software Engineer – Salesforce Developer | Accenture *Hyderabad, India | Jan 2023–Sep 2024*

- Developed enterprise CRM applications using Apex, Triggers, SOQL, Salesforce Flow, and Lightning Components to automate business workflows.
- Built backend automation for case routing, approvals, and business-rule validation using Apex and Salesforce Flow.
- Designed and optimized SOQL queries for data retrieval, validation, reporting, and production troubleshooting.
- Executed data migration, cleansing, validation, and duplicate-prevention workflows using Salesforce Data Loader and validation rules.
- Implemented role-based access using Profiles, Permission Sets, Role Hierarchy, and Organization-Wide Defaults.
- Investigated production data inconsistencies, application defects, and automation failures, coordinating fixes with development and business teams.

SELECTED DATA AND AI PROJECTS

RetailPulse Data Platform | github | *Python, PostgreSQL, S3, AWS Glue/PySpark, Snowflake, dbt, Apache Airflow*

- Built retail ETL/ELT pipelines integrating 6 PostgreSQL source tables through S3 Bronze/Silver layers into Snowflake dimensional models.
- Implemented watermark-based incremental extraction with 5,000-row fetch batches and watermark updates after successful S3 uploads.
- Orchestrated a 10-task Airflow DAG; demonstrated 6 concurrent Glue runs completing in 67–88 seconds with zero retries.
- Developed 6 Gold marts, 43 dbt data tests, and 27 Glue validation checks; verified 10,040 Silver customer records with Athena.

Analyst Agent | github | *Python, FastAPI, Pandas, DuckDB, PostgreSQL, Gemini API*

- Built a two-agent workspace supporting junior-analyst EDA, dataset profiling, human-approved cleanup, and natural-language SQL.
- Implemented 5 cleanup operations and 7 missing-value strategies with before/after row-count, duplicate, and missing-value validation.
- Integrated guarded DuckDB execution and PostgreSQL JSONB staging with 1,000-row insert batches, preserving raw data for reprocessing.
- Added 26 regression test cases across 6 modules covering data quality, SQL restrictions, raw-data preservation, mapping, and reconciliation.

Kitsune Adaptive Threshold NIDS | github | *Python, NumPy, Pandas, KitNET, Matplotlib, psutil*

- Extended Kitsune's autoencoder-based intrusion detector with adaptive RMSE thresholds to respond to changing network traffic.
- Implemented a 2,000-score rolling window and 100-score drift buffer with stability checks and spike rejection.
- Compared 3 configurations on 764,137 Mirai packets each; saved benchmarks show 7-packet detection latency versus 103 for baseline.
- Evaluated sensitivity trade-offs: z=4.0 lowered pre-attack alerts from 949 to 647 versus z=3.5, with latency increasing from 4 to 7 packets.

EDUCATION

M.S. Computer Science | Texas State University | San Marcos, TX *Jan 2025–Dec 2026 expected | GPA 3.90/4.00*
Coursework: Database Theory and Design, Algorithm Design and Analysis, Machine Learning and Applications, Advanced Data Mining

B.Tech. Computer Science | GITAM Deemed University | Visakhapatnam, India *Jun 2018–Jul 2022 | GPA 3.50/4.00*

CERTIFICATIONS

Microsoft Certified: Fabric Data Engineer Associate | Microsoft Certified: SQL AI Developer Associate | NVIDIA: Building AI Multimodal Systems | Salesforce Certified Platform Administrator | Salesforce Certified Platform Developer