Rushi Vasantkumar Parikh

Tempe, Arizona, United States • 623-986-6554 • parikh.rushi.2002@gmail.com • LinkedIn • Github

EDUCATION

Master of Science in Data Science — **August 2024 - May 2026**
Arizona State University, Tempe, United States — GPA: 3.78/4.0

Bachelor of Technology in Computer Engineering — **August 2020 - May 2024**
Gujarat Technological University, Ahmedabad, India — CGPA: 8.5/10.0

SKILLS

Data Engineering & Processing: Apache Spark, Pandas, NumPy, ETL Pipelines, Data Lakes, Streamlit
Programming & Scripting: Python, SQL, JavaScript, Java, TypeScript, C/C++, R, Bash
Data & Cloud Platforms: AWS (S3, EC2, Glue, Athena, Lambda, DynamoDB), Google Cloud, Azure, MySQL, MongoDB
Machine Learning & AI: TensorFlow, PyTorch, scikit-learn, LangChain, FastAPI, Flask
DevOps & Automation: Docker, Kubernetes, Jenkins, Terraform, GitHub Actions, Linux
Version Control: Git, GitHub
Relevant Coursework: Analyzing Big Data, Advanced Database Management Systems, Data Mining, Cloud Computing

PROFESSIONAL EXPERIENCE

NeuroLeap Corp, USA | SDE AI-ML Intern — **May 2025 - August 2025**

- Built Python/OCR/API-based data ingestion pipelines for psychological reports, cutting manual processing time by 75%.
- Designed ETL workflows to clean/validate multi-format data (JSON, CSV, DOCX), processing 1000+ reports/week with 98% accuracy.
- Integrated cloud storage solutions enabling scalable access to 5GB+ of processed lab/test data.
- Collaborated with engineers, data scientists, and backend teams to align pipelines with deployment needs.

SAP Code Unnati, India | AI-ML Intern — **Jan 2024 - April 2024**

- Created data preprocessing and feature engineering pipelines, improving ML training efficiency by 30%.
- Deployed cloud-hosted ML workflows on SAP BTP, reducing deployment time from hours to <15 min.
- Automated data validation for 500K+ records, ensuring 99% error-free ingestion.
- Coordinated with engineers, analysts, and domain experts to align data engineering with testing and integration needs.

PROJECTS

Advanced Transit Data Lake Analytics | Python, AWS

- Developed a scalable cloud-based ETL pipeline for GTFS datasets with automated schema management, feature engineering, and real-time inference dashboard.
- Processed 1M+ transit records weekly on AWS (S3, Glue, Athena, Lambda) with MAE of 24.44s in delay prediction.

Pet Adoption & Care System | T-SQL, SQL Server

- Designed a normalized database (8+ tables) with stored procedures, triggers, and audit tracking for a complete pet adoption and care management system.
- Optimized queries and data constraints to handle 10K+ transactional records with zero referential integrity violations.

Auralith – Intelligent Multi-Modal AI Assistant | Python, LangChain, OpenAI

- Built a multi-modal assistant handling voice, text, and PDF inputs using retrieval-augmented generation and interactive Streamlit UI.
- Processed and indexed 100+ multi-format documents for real-time semantic search with sub-second retrieval latency.

Multi-Modal Route Finding System | Python, Streamlit

- Integrated text, image, and voice-based route finding into a single application with map visualization and route statistics.
- Achieved 95%+ route match accuracy on test maps, enabling end-to-end processing in under 3 seconds per query.

ACHIEVEMENTS & PUBLICATIONS

Published E-Commerce Fraud Detection Using Machine Learning in Springer CCIS 2024.
Selected as one of the top 5 finalists in a university-level hackathon, building a real-time dashboard for streaming data.
Presented research on Optimization Algorithms for Large-Scale Machine Learning at a graduate seminar, receiving recognition for clarity and technical depth.