

Shruti Borkar

+1 (413) 510-7393 | shrutiborkar13@gmail.com | LinkedIn | Github | Portfolio

Data Engineer with 2+ years building production **ETL pipelines** and AI workflows on **Databricks, Spark, Airflow, and Delta Lake** – with hands-on **LangChain, RAG, and Agentic AI** expertise. Delivered 65% reporting reduction, 99.9% forecast accuracy, and 75% ML uplift in enterprise environments.

PROFESSIONAL EXPERIENCE

Five Edc Architects

January 2021 – May 2023

Data Engineer

- Reduced manual reporting by 65% by designing scalable SQL/Databricks ETL workflows and structured data models, improving data warehousing reliability and operational efficiency.
- Reduced RFP delays by 28% by delivering Power BI dashboards with DAX KPIs and presenting data storytelling insights to leadership, accelerating strategic decisions.
- Engineered an AI-powered knowledge graph using SCM data to forecast vendor efficacy, reducing delays by 28%.
- Improved forecast accuracy to 99.9% by automating KPI pipelines in partnership with finance and operations, strengthening enterprise BI capabilities.

Tata Consumer Products

May 2023 – July 2023

Data Science and AI Intern

- Built ETL pipelines with SQL, Python, and Alteryx to process 1M+ sales records with integrated dashboard components.
- Performed data profiling, normalization, and encoding via rule-based validation and statistical techniques to identify and correct data anomalies, improving downstream model reliability.
- Developed collaborative filtering and regression models, boosting product recommendation accuracy by 75%.
- Applied feature engineering and preprocessing to improve model performance and interpretability.
- Created SQL-based data pipelines and deployed AI-driven ThoughtSpot dashboards, reducing reporting time by 45%.

PROJECTS

Netflix Genre Intelligence and Agentic AI Pipeline | Databricks, OpenAI API, LangChain, Power BI, Python, SQL

- Transformed a raw 7,000+ title Netflix dataset using Python (Pandas), SQL, and Excel; normalized multi-country and multi-genre rows, standardized durations, and engineered genre gap metrics for 140+ countries.
- Applied K-Means clustering, built regression-based exploratory insights, and used LLMs (GPT-4 via OpenAI API) to generate country-level recommendations with a Streamlit Q&A chatbot UI.
- Prototyped on Databricks to scale preprocessing and GPT-based recommendation tasks across 140+ country datasets.
- Delivered Power BI dashboard with DAX KPIs to help identify genre gaps and prioritize localized content investments.

Human Behavior Classification via ML | Python, Scikit-learn, Random Forest, Statistical Analysis

- Analyzed the ExtraSensory dataset (60 users, 300K+ samples) to build a multi-label activity classification pipeline using sensor data; achieved 94% accuracy with Random Forest by identifying the top 3 sensor feature groups per label
- Outperformed Logistic Regression baseline by 17% F1-score; validated feature relevance using t-test, Chi-square, ANOVA, and correlation analysis – establishing a reusable statistical QA framework for sensor ML pipelines

TECHNICAL SKILLS

Languages: Python(Pandas, PySpark, NumPy), SQL(MySQL, PostgreSQL, BigQuery), R, DAX, MS Excel(Pivot, Macros)

AI / ML / LLMs: OpenAI API, LangChain, Agentic AI, RAG Pipelines, MLflow, Scikit-learn, NLP, Prompt Engineering

Data Platform: Databricks, Apache Spark, Apache Airflow, ETL Pipelines, Delta Lake, Snowflake, Data Warehousing

Cloud & BI: AWS (S3, Glue), GCP (BigQuery), Azure, Power BI, Tableau, ThoughtSpot, Amazon QuickSight, MATLAB

PUBLICATIONS & CERTIFICATIONS

- **IJCA 2024:** Borkar et al. “Efficacy check of Haralick and Symmetry features for Skin Lesions Classification”
- **Databricks:** AI Agent Fundamentals; SQL Analytics and BI on Databricks
- **Microsoft:** Power BI Data Analyst Specialization Professional Certificate
- **IBM:** Generative AI for Data Analysts Specialization Professional Certificate

EDUCATION

University of Massachusetts Amherst

M.S. in Business Analytics — GPA: 4.0 / 4.0

Amherst, MA

Expected May 2026

NMIMS University

B.Tech in Computer Science and Business Systems — CGPA: 3.57 / 4.0

India

May 2024