

Vaikunth Desai

vdclassifier@gmail.com | +1 (669) 588-6967 | Houston, TX

[linkedin.com/in/vaikunthdesai](https://www.linkedin.com/in/vaikunthdesai) | github.com/vaikunthd | vaikunthdesai.com

SUMMARY

Data Engineer with ~2.5 years building, optimizing, and governing data pipelines for regulated, sensitive financial data. Strong in Python, SQL, and PySpark on Azure Databricks and Data Factory, with hands-on data masking, Unity Catalog governance, performance tuning, and Git-based CI/CD workflows. Statistics and ML background in anomaly detection and forecasting.

EXPERIENCE

AI Data Engineer — KPMG US · Remote *Dec 2025 – Present*

- **Built a reusable data masking function** adopted across the project team to restrict sensitive financial data from unauthorized users, and led a knowledge-transfer (KT) session on the end-to-end workflow.
- **Introduced Unity Catalog metric views** on key tables to standardize metric definitions and enable consistent, governed reporting; authored team documentation to drive adoption over standard views.
- **Applied performance optimization** on large tables, scaling processing from ~5 months to ~5 years of data while cutting query run time by ~30%.
- **Developed and modified PySpark and SQL transformations** in Azure Databricks from business user stories, and built, ran, and maintained pipelines in Azure Data Factory.
- **Delivered changes through Git pull requests within the team's CI/CD workflow**, with automated review and controlled deployment across Databricks environments.

Data Scientist — Capital One · Remote *Jan 2025 – Nov 2025*

- **Wrote PySpark and SQL in Azure Databricks** to clean, prepare, and process data for model training, reducing data-preparation time by ~15%.
- **Built and tested ML models, including Isolation Forest**, to flag anomalies and high-risk transactions, reaching ~90% precision in evaluation.
- **Developed time-series forecasting models (Prophet, ARIMA)** for financial metrics, improving forecast accuracy by ~15%.
- **Supported CI/CD for ML models with GitHub Actions and Docker**, and built Tableau and Power BI dashboards to share results with stakeholders.

Data Analyst — MultiTech Developers · Gujarat, India *May 2020 – Nov 2021*

- **Wrote SQL to extract, clean, and analyze data** for business reporting, with data validation to keep reports accurate and consistent.
- **Built and automated Power BI dashboards** to track KPIs, reducing manual reporting effort by ~30%.
- **Created Excel pivot-table reports** for sales and revenue analysis to support marketing and product teams.

PROJECTS

- **LinkedIn Post Summarizer Chrome Extension** — JavaScript, OpenAI, AWS, LLM: real-time extension summarizing professional posts, backed by a serverless AWS backend, local storage, and an interactive UI.
- **Customer Retention & BI Analysis** — Power BI, Python, DAX, ML: retention models and automated profitability dashboards highlighting where product-segment margins could improve.
- **Walmart Sales Forecasting System** — Python, scikit-learn, Regression: regression models forecasting regional sales across 50+ locations, integrated into Excel-based planning tools.

SKILLS

- **Programming:** Python, SQL, PySpark, R
- **Data Engineering & Cloud:** Azure (Databricks, Data Factory), AWS, GCP, Unity Catalog, ETL pipelines, data masking & governance, performance tuning, CI/CD (GitHub Actions, Docker), Git
- **Machine Learning:** Forecasting (Prophet, ARIMA), Anomaly / Fraud Detection (Isolation Forest), Regression, Classification
- **BI & Databases:** Power BI, Tableau, Excel; PostgreSQL, MySQL

EDUCATION

M.S., Statistics — San Francisco State University, California *Jan 2022 – Dec 2023*

B.E., Computer Engineering — Gujarat Technological University, India *Jul 2017 – May 2021*

CERTIFICATIONS & PUBLICATIONS

- **Certifications:** Google Data Analytics; Snowflake Hands-On Essentials — Data Warehousing Workshop.
- **Publications:** Published research papers on [Big Data](#) and on [COVID detection using CNNs](#) in an international journal.