

Devam Shah

Data Scientist & Analytics Engineer

Brampton, ON • devamshah26@gmail.com • linkedin.com/in/devam-a-shah • github.com/dshah-ds • devam-shah.com

PROFESSIONAL SUMMARY

Data professional with 5+ years of experience turning multi-source data into reliable reporting, forecasts, and business decisions. Most recently owned an end-to-end campaign-to-client data pipeline, from ad campaigns through leads, onboarding, and tax processing, building the dashboards, sales forecasts, and validation checks that kept it trustworthy. Strong in SQL/T-SQL, Power BI, and Excel, with working knowledge of Azure Data Factory, Databricks/PySpark, and Snowflake. Builds machine learning and GenAI projects in Python (regression, tree ensembles, neural networks, NLP, Gemini API), with a habit of catching data leakage and reporting honest results. Brings banking-grade discipline in data quality, root-cause analysis, and documentation.

TECHNICAL SKILLS

Data & Querying: SQL (complex multi-source joins, aggregations, reconciliation), T-SQL, MySQL, SQLite, Excel (advanced formulas, pivot tables, Power Query)

Visualization & Reporting: Power BI (DAX, Power Query), KPI dashboards, Matplotlib, Seaborn, data storytelling for technical and non-technical audiences

Analytics & Machine Learning: Python (pandas, NumPy, scikit-learn, statsmodels, SciPy), regression (OLS, Ridge, Logistic), Decision Tree, Random Forest, KNN, K-Means & hierarchical clustering, PCA, GridSearchCV, ROC-AUC tuning, hypothesis testing, forecasting

Deep Learning, NLP & GenAI: TensorFlow/Keras (ANNs, dropout, L1/L2, batch norm, early stopping), NLTK/VADER, TF-IDF, Google Gemini API, Gradio

Data Platforms & Cloud: Azure Data Factory, Terraform, Databricks/PySpark, Snowflake (working knowledge), Azure Static Web Apps, GitHub Actions CI/CD, Git/GitHub

Operations & ITSM: incident and root-cause analysis, SLA reporting, ServiceNow, ZenDesk, JIRA, Active Directory

PROFESSIONAL EXPERIENCE

Data Specialist (Seasonal Contract – Tax Season) | TaxMobileCanada | Ontario Feb 2026 – May 2026

- Owned the end-to-end data pipeline from ad campaigns to leads, client onboarding, and tax processing; implemented validation checks that identified and resolved data leakage between stages.
- Analyzed targeted advertising data to evaluate lead quality and tie campaign performance to conversion outcomes, supporting targeting and measurement decisions.
- Built and maintained dashboards tracking lead generation, conversion rates, and processing volumes, giving the business owner one view of funnel performance.
- Produced sales forecasts from lead volume and conversion trends to support planning through peak tax season.
- Maintained a centralized client database and documented processes for a planned handover to the business owner at contract end.

Solutions Engineer | OSL Retail Services | Ontario Oct 2024 – Feb 2026

- Investigated and resolved technical issues across retail technology systems, using incident and log data to find root causes and recurring failure patterns.
- Tracked SLA performance and incident trends and reported status to cross-functional stakeholders and clients.
- Coordinated deployments, hardware installations, and integrations with cross-functional teams, and documented procedures for repeatable delivery.

Technical Systems Analyst | TD Bank | Toronto, ON Nov 2020 – May 2024

- Wrote complex SQL and T-SQL queries to extract, join, and reconcile data from multiple enterprise sources for operational reporting and issue investigation.
- Built Power BI and Excel reports tracking operational KPIs and incident trends for business and technology stakeholders.
- Performed data validation and quality checks, traced discrepancies to root cause, and documented fixes to prevent repeat issues in a regulated banking environment.
- Supported data ingestion and transformation workflows using Azure Data Factory and Databricks (PySpark).

SELECTED PROJECTS

Chicago Crime Classification (ANN) | Python, TensorFlow/Keras • github.com/dshah-ds/chicago-crime-ann-classifier

Classified ~400,000 crime records ingested live from the Chicago Open Data API; caught a leaking feature behind a 99.8% score and retrained to an honest 77.4% test accuracy, then translated results into patrol, staffing, and seasonal-planning insights.

Customer Sentiment & Review Automation (GenAI) | Python, NLTK, scikit-learn, Gemini, Gradio • github.com/dshah-ds/customer-sentiment-review-automation

Turned 3,000 product reviews into sentiment scores, segments, and a human-in-the-loop reply assistant; found 45% of reviews had star-rating vs text-sentiment mismatch.

Property Price Prediction (Ames Housing) | Python, statsmodels, scikit-learn • github.com/dshah-ds/ames-housing-price-prediction

Built and compared OLS, Ridge, Decision Tree, Random Forest, and KNN with p-value backward elimination, VIF checks, and GridSearchCV; published the analysis on Medium.

ReadmitRadar: Hospital Readmission Classifier | Python, TensorFlow/Keras • github.com/dshah-ds/readmitradar-hospital-readmission

Built a neural network on 5,000 clinical records and ran 5 documented regularization trials to diagnose overfitting; best 62.6% test accuracy with a 64→32→16 architecture.

CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate (DP-203) — Expired June 2025

EDUCATION

Data Science & Machine Learning Program | Career247 (PwC-endorsed, NASSCOM-certified) | Dec 2025 – Nov 2026

PG Diploma, Software & Information Systems Testing | Fanshawe College, ON | 2017 – 2018

PG Diploma, Project Management | Fanshawe College, ON | 2016 – 2017

B.E., Computer Engineering | Gandhinagar Institute of Technology, India | 2012 – 2016