

VIRAJ SHETTY

Credit Risk Model Governance | Data Scientist
virajvshetty47@gmail.com · +1 4693887873 · Dallas, TX
[LinkedIn](#) · [GitHub](#) · [Portfolio](#)

EXPERIENCE

Apex Informatics Dallas, TX

September 2026 - Present

AI/ML Engineer

- Implemented an **AI-powered text-to-SQL application** translating natural-language requests into operational queries, integrating automated **Hermes regression testing** to validate query accuracy for stakeholder use.
- Streamlined **SQL and ETL pipelines** integrating project, timesheet, client, and financial data from **5+ internal sources**, standardizing datasets for downstream analytics and operational reporting.
- Configured **AI agents** to automate recurring reporting and **data quality validation** using Python and SQL, coordinating repeatable reporting workflows and reducing reporting turnaround by **4 hours**.
- Built an **AI-powered Streamlit report** with **natural-language explanations** of operational metrics, translating trends in resource utilization, SLA performance, and budget variance into insights for operational planning.
- Engineered automated workflows using **Clay, HubSpot CRM, and n8n**, integrating lead enrichment, prospect segmentation, CRM synchronization, and multistep outreach to streamline revenue operations.

Elevate Credit, Dallas, TX

August 2023 - July 2026

Data Scientist II

- Standardized the model development lifecycle by reviewing the model approach, target, and segmentation, and replaced email-based reviews with structured leadership alignment meetings, **reducing turnaround time by 75%**.
- Presented internal & external stakeholders model performance reviews, flagging underperforming models and communicated strategy changes using Copilot meeting notes, retired **2** underperforming models in a single review.
- Drove development of a **RAG-based** document review tool (**LangGraph**) that auto-flagged updates and answered questions related to underwriting documents, cutting review time by **90%** per underwriting document.
- Owned model inventory lifecycle management **10-12** active models, including retirement of deprecated models, delivering real-time status updates to internal stakeholders and bank partners within **24 hours** of any model change.
- Took ownership of coordinating with third-party vendors to validate model replication and compliance with **SR-26-2 & OCC-11-12**, and **facilitated fair lending reviews** to ensure models were free of discriminatory bias.
- Built an **XGBoost** credit risk charge off model using **stratified k-fold cross-validation** and applying reject inference to optimize performance, reducing projected delinquent loans by **\$1M**.
- Automated variable stability (PSI) and performance (AUC, KS, lift) tracking for **10** active models via parameterized Jupyter notebooks and an engineered Snowflake-to-Power BI pipeline, reducing execution time by **95%**
- Designed ad-hoc analytical views for **charge-off benchmarking, Roll-rate analysis, and re-parameterization comparisons**, driving proactive monitoring and pre-rollout lift analysis reducing delinquent population by **60%**.
- Enhanced an automated lookup query that retrieves underwriting data by application ID upon API-flagged fund/no-fund conflicts, leveraging **Claude** for query refinement, cutting manual investigation time by **85%**.

Loopback Analytics, Dallas, TX

June 2022 - December 2022

Data Analyst Intern

- Analyzed and tracked client feasibility requirements using **JIRA** and used **Snowflake** to write stored procedures to optimize execution time and space by **25%**, optimized **Power BI** data structure to reduce report load time by **30%**
- Collaborated with a cross functional team, identified **5 key business metrics**, and conducted Hypothesis (t-tests) on patient data to identify differences in treatment outcomes and improving treatment success rate by **20%**
- Utilized **PySpark** to manage and update NLP pipelines for EHR data and used **Databricks & R** to perform **Data Cleaning & ETL** on socio-economic data from **Census API** reducing the null value in the existing data by **30%**

SKILLS

Programming: R, SQL, Rest-API, Databricks, Git, VSCode, Linux, Python ([Pandas](#), [Matplotlib](#), [Numpy](#)).

Data Science: PyTorch, Keras, Scikit-learn, TensorFlow, XGBoost, LangGraph

Visualization tools: PowerBI, Snowflake, Tableau, Microsoft Excel ([VLOOKUP](#), [Macros](#), [VBA](#), [Index-Match](#))

Cloud: Docker, Azure ([VM](#), [Repository](#), [Storage Management](#)), AWS ([S3](#), [Athena](#), [Redshift](#), [Glue](#))

AI/LLM: Claude, GitHub Copilot, ChatGPT, Perplexity, Databricks Genie AI

PROJECTS

Credit Risk Charge-Off Prediction Model, *Personal Project* [\[Code\]](#) Python | XGBoost

- Built an XGBoost credit risk model on **2.3M+** Lending Club loans, using reject inference (fuzzy augmentation) to incorporate **27M+** declined applications and correct for accept-only sample bias.
- Tuned via stratified 5-fold CV to **0.709** AUC / **0.303** KS; ran segment analysis across grade, DTI, and FICO bands

Fraud Transaction Detection, Wells Fargo, *ML Project* [\[Code\]](#) Tableau Prep | Python | H2O

- Cleaned and prepared **1M** rows of transaction data from Wells-Fargo for analysis by **one-hot encoding** categorical variables and conducted **feature engineering** to identify and select relevant variables
- Implemented **XGBoost** model for fraud detection by using cross-validation techniques such as **grid & random** search to obtain the best hyperparameters, resulting in a robust model with **93%** accuracy

Querycraft: Text-2-SQL, *Work Project* [\[Code\]](#) Python / Flask | SQLite | SQLglot

- Built Codex compatible **Querycraft** using **Python, Flask, SQLite** to generate schema-grounded SQL from natural text, developed a JavaScript workbench with schema exploration, editable queries, and CSV export.
- Designed a **SQLGlot validation pipeline** with user-approved, read-only execution, query deadlines, and row limits which was verified functionality through **92 AI automated tests** and browser regression checks.

Alzheimer's Prediction, *Deep Learning Project*, [\[Code\]](#) TensorFlow | Matplotlib

- Utilized a dataset of brain MRI images, split into training, validation, and test sets with an **80:10:10** ratio.
- Applied image preprocessing techniques to resize images to **128x128** pixels. Employed dropout regularization method to improve the model's ability to generalize and handle potential overfitting.
- Developed a deep learning model using a 2D CNN to accurately classify brain MRI images and detect dementia with 92% accuracy on test set.

EDUCATION

The University of Texas at Dallas, *Dallas, TX*

MS in Information Technology Management (*Business Analytics Specialization*)

August 2021 - May 2023

Relevant coursework: Machine Learning, Advanced Statistics for Data Science, Database Foundations, Big Data

The University of Mumbai, *Mumbai, India*

BE in Computer Engineering

June 2016 - November 2020

Relevant coursework: Analysis of Algorithms, Cloud Computing, Natural Language Processing, Data Warehousing