

Avikshith Yelakonda

+1 (945) 998-3967 | avikshith.y@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Professional Summary

Data Scientist with over 2 years of experience designing scalable machine learning systems, deploying models to production, and automating CI/CD workflows using Python, SQL, and platforms such as MLflow, Airflow, and Databricks. Proficient in building and optimizing models using PyTorch, TensorFlow, and advanced statistical methods across AWS, GCP, and Azure. Adept at translating data-driven insights into resilient, high-impact solutions that deliver measurable business outcomes in dynamic environments.

Skills

Programming & Tools: Python (pandas, NumPy, scikit-learn, PyTorch, TensorFlow), R, SQL, PySpark, Git, Jupyter, VS Code

Machine Learning, Deep Learning & NLP: Supervised and unsupervised learning, regression and classification, clustering, forecasting, transformers, LLM fine-tuning, embeddings, topic modeling, NLP pipelines, prompt engineering, computer vision, hyperparameter tuning, model calibration and validation, feature engineering

Statistical Modeling & Research: EDA, hypothesis testing, probability models, A/B testing, causal inference, optimization, time-series analysis

Data Engineering & Pipelines: ETL and ELT workflows, Airflow orchestration, PySpark and Databricks transformations, data ingestion, preprocessing, validation frameworks, metadata modeling, data wrangling and integration

Cloud & Big Data: Azure, AWS (S3, Lambda, Glue, Redshift), Snowflake, Hadoop ecosystem, distributed compute, scalable training and inference environments

Visualization & Reporting: Tableau, Power BI, Matplotlib, Seaborn, KPI dashboards, performance monitoring, insights communication

Software Engineering & MLOps: Model deployment, reproducible pipelines, version control, CI/CD automation, experiment tracking with MLflow, containerized workflows

Collaboration & Communication: Technical writing, analytical documentation, stakeholder communication, cross-functional collaboration across product, engineering and business teams, data storytelling

Experience

PioneerSoft (Client: Freddie Mac) | *Data Scientist* | USA

July 2025 – Present

- Built ML/DL models using PyTorch/TensorFlow to predict borrower behavior, improving risk classification accuracy by 18%.
- Engineered features and cleaned large-scale loan and borrower datasets with Python and SQL to support credit analytics.
- Designed & evaluated supervised, unsupervised, & DL models using robust metrics to ensure stable, compliant risk outcomes.
- Automated ETL pipelines using Databricks and Snowflake to deliver validated, analytics-ready mortgage datasets.
- Integrated ML models into cloud systems using GitHub, CI/CD pipelines, and version control for reproducible deployments.
- Conducted EDA, statistical testing, and model validation to improve predictive reliability and housing finance alignment.
- Built transformer-based NLP prototypes to extract insights from servicing notes and documents for risk forecasting.
- Partnered with engineering and product teams to productionize ML workflows supporting scalable mortgage risk decisioning.

Yogin | *Data Analyst* | India

May 2021 – May 2023

- Cleaned, merged, and modeled structured datasets using Python, R, and SQL to support analytics pipelines and ML readiness.
- Developed regression, clustering, and classification models to uncover insights and support data-driven operational decisions.
- Automated SQL and Python ETL workflows with Power BI and Tableau, reducing manual reporting effort by 35%.
- Performed EDA, hypothesis testing, and trend modeling to identify key drivers and guide data-backed recommendations.
- Designed dashboards to visualize KPIs, model outputs, and performance trends across multiple business functions.
- Validated data quality using QA scripts, anomaly detection checks, and consistency rules to ensure reliable analytics.
- Collaborated with cross-functional teams to refine requirements and deliver actionable, insight-driven solutions.
- Used Git and Docker to build reproducible analytics environments across local and cloud-based setups.

Projects

Consumer Affairs Lead Conversion Case Study | Python, Scikit-learn, XGBoost, Power BI, Regression

- Built an end-to-end **predictive modeling pipeline** (Logistic Regression, XGBoost) with **ROC AUC ≈ 0.68** , enabling **2.3–2.4x lift in top-decile leads** and delivering actionable insights via **business-ready dashboards**.

Financial Literacy LLM Fine-Tuning Project | Python, PyTorch, Hugging Face, PEFT/LoRA, TinyLlama, OpenAI, Streamlit

- Fine-tuned a **TinyLlama LLM** with **LoRA/PEFT** on curated financial Q&A data, built the full **training–evaluation–chat** pipeline, and deployed an interactive **financial tutor chatbot** providing accurate credit, budgeting, and savings guidance.

Machine Learning & AI with Python Course Projects | Python, NumPy, TensorFlow, Keras, CNN, Transformers

- Built and optimized ML & DL models (Logistic Regression, MLPs, CNNs, Transformers) using NumPy, TensorFlow, and Keras. Applied PCA, feature extraction, adaptive learning, and regularization on diverse datasets, showcasing the full pipeline design. Developed AI agents to solve search & optimization problems using BFS, DFS, A*, Hill Climbing, and Minimax.

University Course Teaching Assistant RAG System | Python, OpenAI, SentenceTransformers, FAISS, TF-IDF, RAG, Docker

- Built a full-stack **multi-tenant RAG platform** where professors upload PDFs/PPTX and students receive **grounded answers** strictly from course materials using a hybrid **TF-IDF + FAISS retriever**, an automated chunking/embeddings pipeline, Streamlit portals for professor/student workflows, and Dockerized deployment.

City Sanitation Multi-Source RAG Assistant | Python, LangChain, LlamaIndex, OpenAI, FAISS, Streamlit, RAG

- Built an **Advanced RAG system** integrating **web, PDF, and CSV data** using **LangChain, LlamaIndex**, and **FAISS** to deliver concise, citation-based answers on city sanitation and recycling services.

Education

Southern Methodist University | *MS in Computer Science - AI/ML* | *GPA: 3.5/4* | Dallas, TX

May 2025