

Sanjay Dilip

+1 (469) 274-5546 | Dallas, TX (Open to relocation)
sanjay.dilip3012@gmail.com | <https://www.linkedin.com/in/sanjaydilip/> | <https://sanjay-dilip.github.io>

Data analyst and ML practitioner in Python, SQL, Power BI, and Azure across predictive modeling, BI reporting, and cloud pipeline design. Built and validated projects on real datasets where the output had to hold up: a credit risk model at 0.94 AUC, supply chain delay prediction at 1.66x lift, and a Kaggle-submitted NLP classifier.

EDUCATION

The University of Texas at Dallas

Aug 2024 - May 2026

Master of Science, Business Analytics & Artificial Intelligence (STEM)

GPA: 3.69

Coursework: Database Foundations, Advanced Statistics, Predictive Analytics, Econometrics, Machine Learning, NLP

Sri Sai Ram Institute of Technology, Chennai, India

Sep 2020 - Apr 2024

Bachelor of Technology, Artificial Intelligence and Data Science (STEM)

SKILLS

- **Data Analysis & Querying:** Python, pandas, NumPy, Excel, data cleaning, data validation, exploratory analysis
- **BI & Reporting:** Power BI, Tableau, IBM Cognos, KPI dashboards, DAX, structured reporting
- **Machine Learning, AI & Modeling:** XGBoost, Random Forest, neural networks, CNNs, NLP, HuggingFace Transformers, forecasting, regression, classification, feature engineering
- **Data Engineering & Tools:** Azure Blob Storage, Azure Data Factory, ETL/ELT pipelines, Parquet, Git, MongoDB
- **Soft Skills:** Stakeholder presentation, cross-functional communication, analytical storytelling

PROJECTS

Levis's RAG (AI Diligence Copilot) | Python, FastAPI, Next.js, Supabase, Gemini, BM25 + Dense Retrieval

- Reached 85.0% recall@10 on a 60-question evaluation set, up from a 40.0% baseline, by diagnosing and re-embedding retrieval-diluted answer chunks in a hybrid BM25 and dense-vector pipeline across 41 SEC filings.
- Deployed a live FastAPI backend and Next.js frontend where every answer traces to a specific SEC filing sentence and one of four evidence tiers based on source reliability.

NBA Win Probability Engine | Python, scikit-learn, Streamlit, Google Cloud

- Reached 0.7357 AUC predicting NBA game winners before tip-off by training on three full regular seasons and testing on the following season as a true future holdout.
- Deployed a live Streamlit dashboard with automated post-game refreshes by building a Google Cloud Run and Cloud Scheduler pipeline that rebuilds predictions without manual intervention.

Credit Risk Modeling with AMEX Data | Python, XGBoost, SHAP, Neural Networks

- Reached 0.94 AUC on held-out test data by engineering 13-month customer behavior features from a 1.1M-record AMEX development sample, aggregating delinquency, spend, payment, and balance variables across 30, 90, and 180-day windows, then comparing XGBoost against a Neural Network that scored 0.92.
- Quantified the default rate tradeoff (4.1% vs 7.8%) between conservative and aggressive approval thresholds using SHAP-informed XGBoost simulations.

Multi Label Emotion Detection | HuggingFace, Decoder Models, vLLM, Weights & Biases

- Reached 0.53 macro F1 on an 11-class multi-label emotion detection task by comparing three decoder fine-tuning approaches, tracked in Weights & Biases.
- Found an instruction-tuned decoder model outperformed a classification-head baseline, then deployed it through a vLLM inference pipeline with deterministic decoding to convert generated text into binary labels across all 11 classes.

PROFESSIONAL EXPERIENCE

MedTourEasy

India

Data Analytics Intern

Oct 2023 – Nov 2023

- Predicted blood donor return likelihood at 0.78 AUC by engineering RFM and time-since-donation features from patient records into a binary classification model.
- Redirected outreach priorities by building IBM Cognos and Excel dashboards that surfaced collection volume as the highest-variance metric; the team lead used this to reorient the donor contact strategy.

Evolve Robot Lab

Chennai, India

AI Intern

Jan 2023 – Feb 2023

- Designed and delivered a 4-week Deep Learning curriculum for 4 students, running from ML fundamentals through neural network architecture to a CNN image classification capstone.
- Achieved stable gyroscopic movement on a drone by calibrating PID parameters across X, Y, and Z axes, verified through remote control testing.

Vaayusastra Aerospace

Chennai, India

IoT Intern

Mar 2022 – Jul 2022

- Built a functional ultrasonic obstacle-detection rover with a two-person team by assembling hardware and writing sensor logic that triggered directional responses on obstacle detection.

LICENSES AND CERTIFICATIONS:

SnowPro Associate (Credential ID: S126888-251211-SOL)