# Anup Sarkar

Data Science & Generative Al Practitioner | Former MSSQL DBA

✉ anupsarkar8106@gmail.com   📞 +91 8106754538   📍 Hyderabad   🔗 Portfolio Link

---

Professional Summary

Data Science & Generative Al Practitioner with a strong foundation in database administration and engineering, transitioning from nearly 3 years of experience as an MSSQL DBA. Skilled in data cleaning, exploratory data analysis, feature engineering, and building machine learning models using Python and Scikit-learn. Hands-on with LLM fine-tuning, prompt engineering, and model deployment using Flask. Proven ability to combine analytical skills with data engineering expertise to deliver end-to-end data solutions.

Certifications

Data Science Certificate ⬈
Cedlearn Tech (NASSCOM)
Mar 2025

SQL Server DBA ⬈
Udemy
March 2022

Education

Bachelor of Engineering (B.E), Computer Science & Engineering,   2016 – 2020
Osmania University   Hyderabad

Projects and Work Experience

Machine Learning Engineer – Semantic Book Recommendation System,   07/2025 – 08/2025
CedLearn-Project   Hyderabad
- Built a semantic book recommender using Kaggle's 7k Books Dataset, enriched with emotion analysis and vector embeddings.
- Applied a DistilRoBERTa emotion classifier to book descriptions and generated embeddings with Sentence-Transformers (MiniLM) stored in Chroma DB.
- Implemented multi-stage filtering (semantic similarity, genre, and emotional tone) to improve personalization and relevance of recommendations.
- Designed and deployed an interactive Gradio dashboard displaying book covers, authors, and summaries.
- Tools & Technologies: Python, Pandas, LangChain, Hugging Face Transformers, Chroma DB, Gradio.

Machine Learning Developer - Sarcasm Detection (Deep Learning, NLP),   06/2025 – 07/2025
Scaler-Project   Hyderabad
- Built a sarcasm classification model on 26K+ headlines using TensorFlow/Keras, achieving 83.05% test accuracy.
- Applied text preprocessing (tokenization, padding, lemmatization, stopword removal) to improve NLP performance.
- Designed a Sequential NN with Embedding & GlobalAveragePooling1D layers for semantic feature extraction.
- Visualized data patterns with Matplotlib & Seaborn to guide model optimization.
- Tech & Tools: Python, TensorFlow, Keras, NLTK, NumPy, Pandas, Matplotlib, Seaborn

Machine Learning Developer - Fraud Detection System, *CedLearn-Project* 05/2025 – 06/2025
Hyderabad
- Performed data cleaning, exploratory data analysis (EDA), and preprocessing on approximately 6 million transaction records, including handling missing values and filtering merchant transactions.
- Engineered new features such as transaction hour, transaction day, and merchant destination flag to improve predictive accuracy.
- Applied label encoding for categorical variables and resolved severe class imbalance using oversampling techniques.
- Trained and fine-tuned a Random Forest Classifier, optimizing hyperparameters to enhance model performance.
- Achieved 92 percent recall for fraud cases, validated using classification reports and performance visualizations.
- Tools & Technologies: Python, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

Machine Learning Engineer - IVF Outcome Predictor, *CedLearn-Project* 02/2025 – 03/2025
Hyderabad
- Developed and implemented a machine learning model to predict IVF pregnancy outcomes using Python and scikit-learn.
- Cleaned and pre-processed clinical and treatment data by handling missing values with forward fill and encoding categorical variables.
- Engineered new features like age bins and IVF treatment classifiers to improve model performance and interpretability.
- Achieved 88% accuracy with a Random Forest model and optimized its performance using GridSearchCV for hyperparameter tuning.
- Tools & Technologies: Python, Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn.

Senior Engineer SQL Server Database administrator, *Microland* 01/2022 – 11/2024
Bengaluru
- Administered and optimized 150+ SQL Server instances and 1100+ databases across production and non-production environments, ensuring security, high availability, and peak performance.
- Automated health checks, maintenance plans, and performance tuning, reducing manual workload and improving system reliability.
- Designed and implemented backup/recovery strategies and high availability solutions (Always-On Availability Groups, Failover Clustering) to achieve near-zero downtime.
- Performed query tuning, index optimization, and performance analysis to enhance application response times and throughput.
- Managed version upgrades (20120162022), migrations, and security configurations to meet compliance and business continuity standards.
- Tools & Technologies: SSMS, SQL Server Agent, T-SQL, SQL Profiler, Extended Events, SQL Configuration Manager, DMVs, Always-On Availability Groups, Failover Cluster Manager

Technical Skills

Programming & ML/DL
Python (Pandas, NumPy, Scikit-learn, TensorFlow, Keras, NLP, LSTM, GEN AI)

Databases & SQL
MSSQL, MySQL, SSMS, SSIS, Performance Tuning, Query Optimizations

Visualization & BI
Power BI (DAX, Power Query), Tableau

Cloud & Tools
Azure Data Studio, AWS S3, SQLWorkbench