# Anup Sarkar

Data Science & Generative Al Practitioner | Former MSSQL DBA

anupsarkar8106@gmail.com +91 8106754538 Hyderabad

Portfolio LinkedIn GitHub Indian

## Professional Summary

Data Science and Generative AI practitioner with a strong foundation in database administration and engineering. Bring over three years of proven MSSQL database administration experience, now focused on applying expertise in data cleaning, exploratory data analysis (EDA), feature engineering, and machine learning model development using Python and scikit-learn. Proficient in fine-tuning large language models (LLMs), prompt engineering, and deploying models through Flask APIs. Demonstrated ability to integrate analytical skills with data engineering to deliver scalable, end-to-end data solutions.

## Certifications

**Oracle Cloud Infrastructure 2025 Certified**
Generative AI Professional

**Data Science Certificate**
Cedlearn Tech (NASSCOM)
Mar 2025

**SQL Server DBA**
Udemy
March 2022

## Professional Experience

**Machine Learning Engineer – AI-Powered Mental Health Support Chatbot,** 10/2025
*Elevate_Labs-Intern* Hyderabad

- AI-powered mental health support chatbot, implementing BlenderBot-based dialogue modeling and end-to-end pipeline orchestration.
- Automated dataset acquisition, cleansing, and stratified splitting into train, validation, and test partitions for robust model training and quantitative evaluation.
- Engineered efficient preprocessing workflows in Python, leveraging text normalization and label management to maximize model generalizability.
- Fine-tuned and evaluated transformer-based conversational agents using Hugging Face Transformers and PyTorch, focusing on domain adaptation for sensitive contexts.
- Developed modular scripts for inference and API deployment, enabling seamless conversion of chat models endpoints.
- Deployed and showcased the chatbot using Gradio on Hugging Face Spaces, ensuring public accessibility and interactive user testing through a live demo link.
- Maintained clear project structure and documentation, enabling reproducibility, code sharing, and effective collaboration.
- **Tools & Technologies**: Python, PyTorch, Hugging Face Transformers, Pandas, Scikit-learn, FastAPI, Gradio, Google Colab

**Machine Learning Engineer – Semantic Book Recommendation System,** 07/2025 – 08/2025
*CedLearn-Intern* Hyderabad

- Built a semantic book recommender using Kaggle's 7k Books dataset, enriching features with emotion detection and semantic embeddings.
- Applied a DistilRoBERTa-based emotion classifier on book descriptions to quantify affective tone and reader sentiment.

- Generated dense vector embeddings using Sentence-Transformers (MiniLM) and stored them in Chroma DB for efficient vector retrieval.
- Designed a multi-stage ranking pipeline combining semantic similarity, genre classification, and emotion-aware filtering for personalized recommendations.
- Developed and deployed an interactive Gradio dashboard on Hugging Face Spaces, enabling users to explore and interact with live book recommendations.
- Referenced framework and methodology from Scaler Academy's Recommender Systems modules, enhancing design quality and scalability.
- Improved recommendation coherence and emotional relevance through embedding space optimization using cosine similarity metrics.
- **Tools & Technologies**: Python, Pandas, LangChain, Hugging Face Transformers, Sentence-Transformers, Chroma DB, Gradio, Hugging Face Spaces

**Machine Learning Developer – Sarcasm Detection (Deep Learning, NLP),** 05/2025 – 06/2025
*Scaler-Referenced* Hyderabad

- Built a sarcasm classification model on 26K+ news headlines using TensorFlow and Keras, achieving 83.05% test accuracy.
- Applied text preprocessing (tokenization, padding, lemmatization, stop-word removal) to enhance model generalization.
- Designed a Sequential Neural Network with Embedding and GlobalAveragePooling1D layers for semantic representation learning.
- Visualized data distributions and model performance using Matplotlib and Seaborn.
- Deployed a live interactive demo on Hugging Face Spaces using Gradio, allowing users to test headline sarcasm predictions in real time.
- **Tools & Technologies:** Python, TensorFlow, Keras, NLTK, NumPy, Pandas, Matplotlib, Seaborn, Gradio, Hugging Face Spaces

**Machine Learning Developer - Fraud Detection System,** *CedLearn-intern* 03/2025 – 04/2025
Hyderabad

- Performed data cleaning, EDA, and preprocessing on ~6 million transaction records, handling missing values and filtering merchant transactions.
- Engineered transaction-based features such as hour, day, and merchant destination flag to improve predictive accuracy.
- Applied label encoding and addressed class imbalance using oversampling techniques.
- Trained and fine-tuned a Random Forest Classifier, optimizing hyperparameters for improved precision and recall.
- Achieved 92% recall for fraud cases, validated through classification reports and visual performance metrics.
- **Tools & Technologies:** Python, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn.

**Machine Learning Engineer - IVF Outcome Predictor,** *CedLearn-intern* 02/2025
Hyderabad

- Developed and implemented a predictive machine learning model for IVF pregnancy outcomes using Python and Scikit-learn.
- Cleaned and preprocessed complex clinical and treatment datasets by handling missing values (forward fill) and encoding categorical variables.
- Engineered new features such as age bins and IVF treatment classifiers to boost model interpretability and predictive performance.
- Achieved 88% accuracy using a Random Forest Classifier, optimized through GridSearchCV for hyperparameter tuning.
- **Tools & Technologies:** Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn.

**Senior Engineer – SQL Server Database Administrator,** *Microland LTM* 01/2022 – 11/2024
Bengaluru

- Administered and optimized 150+ SQL Server , managing 1,000+ databases across production and non-production environments for Waste Management (US), ensuring high availability, strong security, and performance.
- Automated health checks, routine maintenance, and tuning tasks with PowerShell and SQL Agent, reducing manual workload by 60% and enabling fast, reliable incident response.
- Designed and implemented robust backup, restore, and disaster recovery strategies (Always-On Availability Groups, Log Shipping), achieving 99.9% uptime and near-zero data loss.
- Drove advanced query tuning, index optimization, and proactive statistics management to deliver up to 40% performance improvement in mission-critical systems.
- Orchestrated SQL Server version upgrades (2016–2019), migrations, and patch rollouts, ensuring enterprise compliance, business continuity, and minimal downtime.
- Built and maintained SSRS and custom monitoring dashboards, supporting proactive health, SLA reporting, and capacity planning.
- Enforced database security with RBAC, auditing, and encryption, meeting all corporate and regulatory requirements.
- Collaborated closely with US client teams to manage tickets, resolve incidents, and document SOPs — recognized for operational reliability and continuous optimization.
- **Tools & Technologies:** Microsoft SQL Server (2012/2016/2019), MySQL, AWS RDS, SSMS, SQL Server Agent, T-SQL, Profiler, Replication, Always-On Availability Groups, SQL Configuration Manager, DMVs, Failover Cluster Manager.

## Technical Skills

**Programming & Machine Learning** — Competent
- **Languages & Libraries:** Python (Pandas, NumPy, scikit-learn, TensorFlow, Keras)
- **Concepts:** Machine Learning, Deep Learning, NLP, Generative AI, LSTM, Prompt Engineering
- **Model Development:** Feature Engineering, EDA, Model Evaluation, LLM Fine-Tuning

**Visualization & Business Intelligence** — Competent
- **Tools**: Power BI (DAX, Power Query), Tableau
- **Skills**: Dashboard Design, Data Storytelling

**Databases & SQL** — Proficient
- **Technologies**: MSSQL, MySQL, SSMS, SSIS
- **Expertise**: Performance Tuning, Query Optimization, Stored Procedures

**Cloud Platforms & Tools** — Amateur
- **Platforms**: AWS, Azure
- **Related Tools**: Azure Data Studio, AWS S3, SQL Workbench, Flask (for Model Deployment)

## Education

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

**Board of Intermediate, (MPC),** *Narayan Junior College, Hyderabad* 2016
Hyderabad

**Board of Secondary Education,** *St. Marry's High School* 2014
Bellampally