

ANUSREE R

Data Science · Machine Learning · AI Engineering

[linkedin.com/in/anusree-r](https://www.linkedin.com/in/anusree-r) | github.com/Anusree-R | anusreeraghu21@gmail.com | +91 9074435579

PROFESSIONAL SUMMARY

Results-oriented aspiring Data Scientist with a B.Tech CSE (Hons. Data Science and Data Engineering) background and hands-on experience building data-driven and AI-based projects, including a Data Scientist internship at TransOrg Analytics. Proficient in Python, SQL, data preprocessing, exploratory data analysis, visualization, feature engineering, and machine learning. Skilled in applying analytical and predictive techniques to real-world problems and building end-to-end solutions — from data preparation and model development to evaluation and deployment. Eager to leverage data science, machine learning, and AI engineering skills to create impactful business and technical solutions.

WORK EXPERIENCE

TransOrg Analytics — Data Scientist Intern

May 2025 – May 2026

- Preprocessed and cleaned 7+ datasets with 50k+ records for predictive modeling by removing irrelevant, non-numerical, noisy, and redundant columns, handling data inconsistencies, and preparing model-ready datasets for machine learning tasks.
- Contributed to predictive modeling workflows through data preparation, feature selection, model development, and evaluation, while building 5+ interactive Power BI and Tableau dashboards to visualize KPIs, trends, and actionable business insights.
- Designed a 13-node Generative AI agentic pipeline using LangGraph and Gemini for an AI-powered travel itinerary planner, integrating Tavily and 3 external APIs to generate budget-aware recommendations across 5 spending tiers (₹10K–₹5L).

PROJECTS

AI Meal Planner & Grocery List Generator

Aug 2026 – Present

- Architected a 6-agent LangGraph pipeline (Planning, Recipe, Nutrition, Budget, Grocery, Validation) with an automatic retry loop that re-generates a plan against specific constraint violations — reducing generation latency by parallelizing 7 concurrent LLM calls (one per day) instead of a single sequential 21-recipe call.
- Built a secure FastAPI backend with JWT authentication (access/refresh token rotation) and object-level authorization, validated by 21 automated tests including a dedicated cross-user data-isolation suite.
- Implemented a deterministic validation layer that checks every AI-generated meal plan against allergy, dietary, and calorie constraints before it reaches the user — catching constraint violations the LLM itself introduced, rather than trusting model output directly.

Customer Support Ticket Triage & Sentiment Analysis

Jul 2026 – Aug 2026

- Engineered an end-to-end NLP classification pipeline using Python, scikit-learn, TF-IDF (5,000 features, 1–2 grams), and Logistic Regression, achieving 87.52% accuracy and 0.8417 weighted F1-score on 2,964 real-world Twitter customer-support tweets across 5 ticket categories: billing, technical, refund, account, and shipping.
- Developed a production-ready dual-interface inference system with Streamlit and Flask REST API (/predict), delivering ticket category, confidence score, and urgency flag with <50ms per-prediction latency; implemented persisted ML artifacts, reproducible preprocessing, and fallback error handling.
- Created a modular ML architecture separating preprocessing, training, and inference components, and documented model evaluation, per-category performance, and deployment workflows in a comprehensive technical report.

Spotify Genre Predictor

Jul 2025 – Nov 2025

- Constructed an end-to-end ML genre classification pipeline using Python, scikit-learn, and Random Forest, processing 114,000 Spotify tracks and consolidating 114 sub-genres into 13 parent genres using 12 audio features such as energy, danceability, acousticness, and tempo.
- Performed EDA and feature engineering, addressing class imbalance with SMOTE and class_weight='balanced'; identified key predictors including instrumentalness, acousticness, and speechiness, with an energy–loudness correlation of 0.76.
- Refactored the model into a modular MLOps pipeline with separate ingestion, transformation, training, prediction, logging, and exception-handling components; reduced the model from 909MB to 93MB by sampling 20,000 tracks to fit GitHub's 100MB limit, achieving 75% accuracy across 13 genres.

SKILLS

Programming & Databases: Python, SQL

Machine Learning & AI: Machine Learning, Feature Engineering, Model Evaluation, Supervised & Unsupervised Learning, Agentic AI, LangGraph

Data Analysis & Visualization: Exploratory Data Analysis, Power BI, Tableau, Excel

Libraries & Frameworks: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Streamlit, FastAPI

Dev Tools & Platforms: Git, GitHub, Jupyter Notebook, Google Colab, VS Code

GenAI Productivity: ChatGPT, Claude, GitHub Copilot

Soft Skills: Problem-Solving, Communication

ACHIEVEMENTS

- Secured ₹1,00,000 seed funding for early-stage startup development, demonstrating product vision to investors (May 2024)

CERTIFICATIONS

SQL (Advanced) - Hackerrank

Jul 2026

Oracle Certified Foundations Associate — Agentic AI

Aug 2026

EDUCATION

Lovely Professional University — Phagwara, Punjab

Aug 2022 – May 2026

B.Tech CSE (Hons.) Data Science and Data Engineering | CGPA: 7.17