# Pranav Gujjar

Machine Learning Engineer | Data Scientist
Bangalore, India | +91-9008913366 | mr.pranavgujjar@gmail.com
LinkedIn: linkedin.com/in/pranav-b-gujjar | GitHub: github.com/Gujjar-Pranav |
Portfolio: Live ML Systems & Case Studies
**Visa:** India (Citizen) | Open to UK / EU / US roles (sponsorship as per role)

## Professional Summary

Machine Learning Engineer with an MSc in Data Science (Distinction) and hands-on experience designing, evaluating, and deploying production-style ML systems. Strong background in translating structured and unstructured data into actionable insights through robust modeling, explainability, and deployment-focused engineering. Experience spans NLP, speech processing, computer vision, and time-series forecasting, with emphasis on reproducibility, interpretability, and real-world evaluation.

## Core Skills

- **Languages:** Python, SQL, PySpark, R, MATLAB
- **Machine Learning:** Supervised & Unsupervised Learning, Classification, Regression, Clustering, Feature Engineering, Feature Selection, Cross-Validation, Hyperparameter Tuning, ROC-AUC, Precision-Recall, F1-score, MAE, RMSE
- **Deep Learning:** Neural Networks, CNNs, RNN, LSTM, GRU, TensorFlow, PyTorch
- **Computer Vision:** Image Preprocessing, CNN-based Models, U-Net, Dense U-Net, Image Segmentation, Data Augmentation, CLAHE, Patch-Based Learning
- **NLP & Speech:** TF-IDF, Linear SVM, Rule-Based NLP, Whisper ASR
- **Time Series:** ARIMA, Exponential Smoothing, RNN-based Forecasting
- **MLOps / Deployment:** FastAPI, REST APIs, Streamlit, Docker, CI/CD (GitHub Actions), Model Persistence
- **Data & Viz:** PostgreSQL, MongoDB, Tableau, Power BI, Matplotlib

## Work Experience

**Machine Learning Engineer (Freelance)** **Aug 2025 – Present**

- Designed and delivered end-to-end ML systems covering data ingestion, feature engineering, model training, evaluation, and deployment.
- Deployed inference services using FastAPI with Streamlit dashboards for predictions, explainability, and confidence-based decision support.
- Ensured reproducibility through persisted preprocessing artifacts, model versioning, and training–inference parity.

**Data Science Intern** **Jun 2022 – Dec 2022**

Vertexblue Pvt Ltd, India

- Improved forecasting accuracy by **15%** using Python- and SQL-based predictive models.
- Delivered analytics contributing to **10%+** operational cost reduction and reduced manual processing by **30%**.

## Deployed ML Systems (End-to-End)

*Live demos, architecture notes, and source code available at: pranav-gujjar-portfolio.vercel.app*

**ReviewSense AI – Sentiment Intelligence Platform**

- Designed and deployed a full-stack sentiment analysis system evaluated on **293 real-world customer reviews**, benchmarking multiple NLP approaches.
- Selected **TF-IDF + Linear SVM**, achieving **85.55% accuracy** and **0.853 F1-score**, with calibrated probability outputs for risk-aware decision-making.

- Delivered interactive dashboards visualizing sentiment distribution, confidence thresholds, and operational exposure.

**Glass Identification – ML Classification System**

- Built a reproducible ML pipeline including data validation, winsorization, feature engineering, and SMOTE-based class imbalance correction.
- Benchmarked ensemble models and implemented a stacking ensemble achieving **90.70% test accuracy**.
- Deployed the system via FastAPI with persisted preprocessing artifacts and a Streamlit-based inference UI.

**Diabetes Risk Assessment – Clinical Decision Support**

- Developed an interpretable Logistic Regression model on a **768-record clinical dataset**, achieving **ROC-AUC 0.813**.
- Implemented probability-based risk stratification with coefficient-level feature explanations.
- Delivered a web-based decision support tool with audit-friendly outputs and clinical disclaimers.

**Intelligent Task Miner – Audio-to-Task AI**

- Built an offline, privacy-preserving audio intelligence pipeline using local Whisper ASR and deterministic rule-based NLP.
- Converted unstructured meeting audio into structured task data including assignees, priorities, deadlines, and dependencies.
- Exported normalized JSON outputs for workflow automation and enterprise integration.

## Applied ML Studies

**Recommendation System – Content-Based Filtering**

- Built a content-based recommender using cosine similarity over sparse categorical and numerical features, processing **12,294 items**.
- Addressed sparsity, feature scaling, and cold-start challenges common in production recommender systems.

**Neural Networks & Sequence Modeling**

- Implemented feedforward neural networks from scratch to study backpropagation and optimization dynamics.
- Built and evaluated RNN, LSTM, and GRU architectures, analyzing convergence and long-term dependency handling.

**Time Series Forecasting – Production & Financial Data**

- Built forecasting pipelines on seasonal production (**168 monthly**) and FX data (**7,588 daily**).
- Implemented ARIMA and RNN-based sequence models with sliding-window evaluation and residual diagnostics.

**Clustering & Unsupervised Learning**

- Applied K-Means, Hierarchical Clustering, and Gaussian Mixture Models for latent structure discovery.
- Evaluated clustering quality using Silhouette Score and Davies–Bouldin Index.

## Academic Projects

**Retinal Image Segmentation for Cardiovascular Diagnostics (MSc Dissertation)**

- Developed and evaluated U-Net and Dense U-Net models on the DRIVE dataset, achieving **97.34% ROC-AUC**.
- Implemented medical image preprocessing including CLAHE, patch-based learning, and augmentation.

**AI Chatbot for UK Train Ticketing**

- Built an NLP-based conversational system for train travel queries using intent classification and entity extraction.
- Engineered features from **400k+ time-series records** and benchmarked multiple ML models.

## Education

**MSc in Data Science (Distinction)** **Sept 2023 – Sept 2024**
University of East Anglia, UK

## Languages

English (Advanced), Hindi (Advanced), Gujarati (Native)