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
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 building end-to-end, production-ready ML systems. Skilled in translating structured and unstructured data into actionable insights through robust modeling, explainability, and deployment-focused engineering. Experienced across NLP, speech processing, computer vision, and time-series forecasting, with a strong emphasis on real-world evaluation and reproducibility.

Core Skills

- **Languages:** Python, SQL, PySpark, R, MATLAB
- **Machine Learning:** Supervised & Unsupervised Learning, Classification, Regression, Clustering, Feature Engineering, Feature Selection, Cross-Validation, Hyperparameter Tuning (Grid Search, Random Search), ROC-AUC, Precision-Recall, F1-score, Accuracy, MAE, RMSE, MSE
- **Deep Learning:** Neural Networks, CNNs, RNN, LSTM, GRU, Backpropagation, Regularization, Dropout, Batch Normalization, TensorFlow, PyTorch
- **Computer Vision:** Image Preprocessing, CNN-based Models, U-Net, Dense U-Net, Image Segmentation, Feature Extraction, Data Augmentation, CLAHE, Patch-Based Learning, Evaluation (IoU, Dice, Sensitivity, Specificity)
- **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

- Built end-to-end ML systems spanning data ingestion, feature engineering, model training, evaluation, and deployment.
- Deployed REST-based inference services using FastAPI with Streamlit dashboards for predictions, explainability, and confidence analysis.
- Ensured reproducibility through persisted 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.
- Generated analytic contributing to **10%+** operational cost reduction and reduced manual processing by **30%**.

Deployed ML Systems (End-to-End)

ReviewSense AI – Sentiment Intelligence Platform

- Built a production-ready sentiment analysis system evaluated on **293 real customer reviews**, benchmarking multiple NLP approaches.
- Selected **TF-IDF + Linear SVM**, achieving **85.55% accuracy** and **0.853 F1-score**, with calibrated probabilities enabling automated and human-in-the-loop moderation.
- Delivered executive dashboards showing sentiment distribution, confidence bands, and operational risk exposure.

Glass Identification – ML Classification System

- Built an end-to-end ML pipeline with data cleaning, winsorization, feature engineering, and SMOTE-based class imbalance handling.

- Benchmarked ensemble models (RF, Gradient Boosting, Bagging, AdaBoost) and selected a stacking ensemble achieving **90.70% test accuracy**, outperforming tuned RF (**86.05%**).
- Deployed the model via FastAPI with persisted preprocessing artifacts and a Streamlit interface for real-time inference.

Diabetes Risk Assessment – Clinical Decision Support

- Built an explainable Logistic Regression model on a **768-row, 9-feature healthcare dataset**, achieving **ROC-AUC 0.813** and **70.8% accuracy**.
- Implemented probability-based risk stratification with coefficient-driven feature explanations and delivered a web application with patient history tracking and clinical disclaimers.

Intelligent Task Miner – Audio-to-Task AI

- Built a fully offline pipeline using local Whisper ASR and rule-based NLP to convert meeting audio into structured task data.
- Extracted assignee's, priorities, deadlines, dependencies, and contextual reasoning, exporting deterministic JSON for workflow integration.

Applied ML Studies (Modeling & Evaluation)

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 real-world 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, vanishing gradients, and long-term dependency handling.

Time Series Forecasting – Production & Financial Data

- Built forecasting pipelines on seasonal production (**168 monthly**) and USD–AUD exchange rate data (**7,588 daily**), applying stationarity testing (ADF), decomposition, and differencing.
- Implemented ARIMA and RNN-based sequence models (LSTM, GRU) with sliding-window forecasting and evaluated robustness using residual diagnostics and out-of-sample behavior.

Clustering & Unsupervised Learning

- Applied K-Means, Hierarchical Clustering, and Gaussian Mixture Models to discover latent structure in unlabeled data.
- Evaluated cluster quality using Silhouette Score and Davies–Bouldin Index, and interpreted clusters for segmentation insights.

Academic Projects

Retinal Image Segmentation for Cardiovascular Diagnostics (MSc Dissertation)

- Developed and evaluated U-Net and Dense U-Net models for retinal vessel segmentation on the DRIVE dataset, achieving **97.34% ROC-AUC** and **95.71% accuracy**.
- Implemented medical image preprocessing including CLAHE, green-channel extraction, patch-based learning, and augmentation; evaluated using Dice, IoU, Sensitivity, and Specificity.

AI Chatbot for UK Train Ticketing

- Built an NLP-based conversational system for train travel queries (pricing, delays, schedules) using intent classification and entity extraction.
- Engineered features from **400k+ time-series records** and benchmarked multiple models (Linear, RF, Gradient Boosting, KNN) with modular, explainable pipelines.

Education

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

Languages

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