Rihaz Raheem

Alappuzha, Kerala, India | +918547570698 | rihasraheem@gmail.com| GitHub| LinkedIn

Data Scientist with over 4 years of experience in data-driven roles, machine learning NLP, Retrieval-Augmented Generation (RAG), and Computer Vision. Specialized in retrieval optimization and end-to-end ML pipelines from data ingestion to lightweight deployment.

KEY SKILLS

Machine learning : Semantic Search, NLP, RAG, Feature Engineering, Model Evaluation

LLMs & Retrieval: SBERT, FAISS, BM25, LlamaIndex, ChromaDB, RAGAS

Tools & Libraries: Python, TensorFlow, Scikit-learn, FastAPI, SQL(SQLAlchemy), Streamlit.

Computer Vision : YOLOv8, ByteTrack, OpenCV, Deepsort

Analytics skills: Exploratory Data Analysis (EDA), Pandas, NumPy, Advanced Excel, Google Colab.

WORK EXPERIENCE

Scipy Technologies • Remote • 04/2025 - Present

Data Scientist

• Achieved 90%+ faithfulness and context recall in RAG pipelines by implementing RAGAS-based validation and systematic retrieval tuning.

• Improved document retrieval quality by identifying and eliminating near-duplicate content across 400+ PDFs using TF-IDF and SBERT embeddings, leading to more consistent grounding and reduced retrieval noise.

• Designed and validated a CPU-only real-time vision pipeline under edge deployment constraints, achieving ~8.9 FPS while preserving stable multi-person tracking in live streams.

• Diagnosed data leakage and class imbalance during exploratory analysis and introduced corrective feature engineering strategies, improving model robustness as measured by Precision, Recall, and F1-score.

District Industries Centre • Alappuzha • 04/2022 - 03/2025

Enterprise Development Executive

• Analyzed a ~10,000-row enterprise dataset to support identification, evaluation, and onboarding of 500+ enterprises, contributing to full target achievement across multiple development schemes

• Built Excel-based dashboards (Pivot Tables, formulas) to track enterprise status, fund utilization, and scheme-wise performance for district-level monitoring.

SMEG Edulabs • 10/2020 - 03/2022

Academic Assistant

• Designed and implemented a facial recognition-based attendance system using OpenCV, enabling automated image capture and CSV-based attendance logging for classroom environments.

• Built a SMS spam detection model using TF-IDF and Scikit-learn, achieving ~96% precision in filtering malicious messages during offline evaluation.

- Mentored students on academic and mini-projects, guiding data preprocessing, EDA, logic building, and result interpretation using Python.

EduBridge India • 01/2020 - 06/2020

Trainer

- Trained 100+ students in data analysis and Excel-based numerical reasoning

- Delivered hands-on sessions using VLOOKUP, PivotTables, and conditional formulas.

EDUCATION

M.Tech, Signal Processing
Government Engineering College Barton Hill, Thiruvananthapuram (2017 – 2019)
B.Tech, Electronics and Communication Engineering
College of Engineering and Management Punnapra, Alappuzha (2013 – 2017)

PROJECTS

RAG-Based Document Assistant: Architected a RAG pipeline using LlamaIndex, SBERT embeddings, and ChromaDB, combining BM25 and vector retrieval with reciprocal reranking achieving 90-100% context recall and 83–100% retrieval precision, with fully grounded responses validated via RAGAS metrics on a 40-query domain test set.
AI Resume Auditor: Built a modular AI resume auditor using SBERT, TF-IDF, and LLaMA-3 for high-precision matching; featured a PII-redacting security layer, Singleton-optimized resource management, and a secure dual-portal interface for automated bulk candidate evaluation via Streamlit.
Age & Gender Detection: Developed a lightweight CNN with MobileNetV2 transfer learning optimized for edge deployment, reducing age MAE from 5.1 to 4.7 years and increasing gender prediction accuracy from 69% to 88%.
AI-Based Person Detection: Engineered a real-time person detection and tracking using YOLOv8 and ByteTrack for multi-person ID tagging and tracking.

CERTIFICATIONS

- PG Certificate in Entrepreneurship: Small Business Management – Digital University Kerala
- Business Analytics with Excel – Simplilearn
- Data Science and real-time project training on Learn to Build a Real Time Application for Emotional Detector – Null Class

LANGUAGE

- English (Excellent)
- Malayalam (Mother Tongue)