Bharatkumar Gungoman

Mumbai | bharatgungoman2020@gmail.com | 7058583117 | Linkedin | Github

Summary

Full-stack & AI engineer. I've shipped across distributed systems, picked up the internship to see how production really works. Now deep in FOSS, contributing back and building in public.

Education

Bachelor of Engineering in Information Technology, **APSIT** Sept 2023 – May 2027

Experience

SDE Intern – PassionBytes, Thane | **Experience Letter** Dec 2025 – Feb 2026

Tech Stack: Apache Spark, Splink, Python, Flask, React.js, Kubernetes, Docker, Trino, MinIO, MongoDB, PostgreSQL, Git

- Refactored a production data deduplication microservice, migrating from a broken monolithic codebase to a modular architecture with a Flask backend and modern React frontend.
- Fixed and rebuilt the non-functional record linkage pipeline powered by Splink, implemented job orchestration and improving deduplication accuracy via optimized EM training and blocking rules.
- Migrated Apache Spark execution from local mode to a distributed Kubernetes cluster, implementing spark-submit client mode with backend-driven drivers and distributed executors.
- Built a Flask-based service layer with asynchronous job orchestration, progress tracking, and integrations with Apache Spark, Trino, and MinIO with Keycloak authentication.
- Containerized the entire application with Docker and successfully integrated the microservice into the main Zigma Data product.

Projects

1. InferX-ML - End-to-End AutoML Platform for Training and Inference

Technologies: Python, React, Flask, Celery, Redis, PostgreSQL, MinIO, Docker

- Built a full-stack no-code AutoML platform for technical and non-technical users to upload datasets and automatically train, evaluate, and compare 10+ ML algorithms across tabular, time-series, and image data types.
- Integrated Groq/Gemini LLMs to auto-analyze uploaded datasets, detect problem type, recommend target columns, and suggest training strategies and eliminating the need for any manual ML configuration.
- Implemented asynchronous model training using Celery + Redis with SHAP-based explainability, and exported a structured ZIP artifact bundling all model files, preprocessing pipeline, evaluation metrics, and schema ready for direct integration into external pipelines.
- Delivered a post-training interactive Streamlit prediction UI for real-time inference, containerized the entire platform across 6 services via Docker Compose for consistent and reproducible deployment.

2. Research Paper Generator & AI Detection System

- Built an AI-powered research platform that generates papers from user topics (via Gemini/Groq) and verifies uploaded PDF/LaTeX documents for originality with explainable detection scores.

Skills

Languages: Python, TypeScript, JavaScript
Frameworks & Libraries: Flask, FastAPI, LangChain, NumPy, Pandas, Matplotlib, Seaborn, React.js, Express.js
Distributed Systems & Data: Apache Spark (PySpark, Spark SQL), Trino, Splink
GenAI & ML: LangChain, RAG Pipelines, LLM Integrations
Databases : PostgreSQL, MongoDB, Redis, MinIO (S3), VectorDB (Pinecone)
Cloud & DevOps: Docker, Kubernetes, AWS, Git

Achievement and Certifications

1. Hackstreak 1.0 **(Winner)**
2. Tech Head (Dev/DB), **Google Developer Group (GDG)**
3. Google Cloud Study Jam | Google GenAI Study Jam, Participant