

SUMMARY

Results-driven Machine Learning Engineer and M.Tech student specializing in end-to-end ML pipelines, NLP, and large-scale data processing. Proven track record of engineering high-performance classification models over 25M+ record datasets and deploying robust retrieval-augmented generation (RAG) systems. Adept at rapid prototyping, data pipeline optimization, and rigorous ML code evaluation to bridge the gap between research and production.

EXPERIENCE

Zeronsec India Pvt. Ltd.

May 2024 – July 2024

Machine Learning Intern

- Architected an Elastic Stack ingestion pipeline to transform high-volume, unstructured logs into queryable, structured data for downstream analysis.
- Developed automated Python preprocessing scripts utilizing pandas and regex to normalize datasets, significantly improving data quality and consistency.
- Collaborated with cross-functional engineering teams to seamlessly integrate processed data pipelines into core analytics workflows and real-time dashboards.

PROJECTS

Business Entity Resolution (Amazon ML Challenge 2026)

- Stack: Python, DuckDB, XGBoost, pandas
- Built an offline entity-resolution pipeline to process and reconcile approximately 25 million records across three disparate data sources.
- Leveraged disk-backed DuckDB and multi-key blocking strategies to generate over 34 million candidate pairs, achieving a 77.7% recall rate.
- Engineered pairwise similarity features and trained an XGBoost classifier, securing a 0.82 macro F0.5 score on held-out validation data.

PaperTrail – Citation-Grounded Q&A over ML Research Papers

- Stack: Python, LangChain, FAISS, BM25, Hugging Face, LLM APIs
- Engineered a retrieval-augmented generation (RAG) pipeline to answer complex queries over ML research corpora, returning highly accurate answers with precise inline source citations.
- Fused dense embedding retrieval (sentence-transformers, FAISS) with BM25 keyword search, optimizing hybrid rankings to capture both semantic concepts and exact-term queries.
- Tuned chunk sizing and overlap to maximize answer faithfulness, implementing rigorous evaluation protocols against retrieved passages to eliminate unsupported model hallucinations.

Voice Bridge (Smart India Hackathon 2024, Finalist)

- Stack: Flutter (Dart), Text-to-Speech (TTS) Models
- Curated and processed diverse multilingual audio datasets, building speech-to-text transformation pipelines for real-time inference.
- Integrated the ML backend into a highly accessible mobile application designed to facilitate seamless communication for the deaf and mute community.

EDUCATION

IITM Gwalior

2026 – 2028

Masters of Technology in Information Technology

Relevant Coursework: Machine Learning Techniques, Parallel & Distributed Computing, Mathematical Foundations of Data Science, Algorithms

SRM University

2022 – 2026

Bachelor of Technology in Computer Science and Engineering (Big Data Analytics)

CGPA: 9.34

TECHNICAL SKILLS

- **Machine Learning & AI:** Python, NumPy, pandas, scikit-learn, XGBoost, Hugging Face, LangChain, NLP, RAG, Feature Engineering
- **Data & Databases:** SQL, DuckDB, MySQL, PostgreSQL, MongoDB, Elastic Stack, ETL Pipelines
- **Languages & Web:** C/C++, Java, JavaScript (React, Next.js, Express)
- **DevOps & Tools:** Git, GitHub, Docker, Linux, AWS, Flask, Vercel