

Manjit Singh

Applied ML & Data Analyst | DS/AI Practitioner

+91 6900054576
itsmanjit20@gmail.com
Tezpur, Assam, IN
[Portfolio](#)
[LinkedIn](#)
[Github](#)

SUMMARY

Results-driven **Data analyst and ML practitioner** with a strong focus on end-to-end **ML** and **NLP** solutions, utilizing **Python** and tools like **Scikit-learn** and **TensorFlow**. Led data engineering for the ONDC program, consolidating datasets for **25k+** sellers and enhancing reporting efficiency through **automated dashboards**. Engineered innovative **RAG** systems and **sentiment-based** recommendation models, demonstrating a commitment to applying **AI** for social good.

KEY SKILLS

- **Data Science** • **Machine Learning**
- **Deep Learning** • **Statistical Modeling**
- **Exploratory Data Analysis (EDA)**
- **Feature Engineering** • **Model Evaluation**
- **Natural Language Processing (NLP)**
- **Transformers & LLMs**
- **Retrieval-Augmented Generation (RAG)**
- **Vector Databases (FAISS, ChromaDB)**

TECHNICAL SKILLS

Programming: Python, SQL
ML Libraries: Scikit-learn, TensorFlow, PyTorch, Keras, Pandas, NumPy
NLP Tools: NLTK, Transformers (Hugging Face), LangChain
Tools: Power BI, MS Excel
Databases: MySQL, PostgreSQL
Deployment: Streamlit, Flask, Docker
Dev Tools: Git, Jupyter Notebook, VS Code

CERTIFICATIONS

- Fundamentals of LLMs | Hugging Face
- LangChain for LLM Application Development | DeepLearning.AI
- Power BI Essential Training | LinkedIn
- Python for Machine Learning | Great Learning
- SQL for Data Analysis | Udacity
- ChatGPT Prompt Engineering for Developers | DeepLearning.AI

EDUCATION

Executive PG Program in Data Science and AI Oct '23 - Jan '25
IIIT Bangalore Bengaluru, IN
• Secured: 4.00/4.00

Bachelor of Commerce - B. Com Jul '20 - Jun '23
Gauhati University Guwahati, Assam, IN
• Secured: 6.89 CGPA

KEY PROJECTS

[HelpMate AI – Retrieval-Augmented Generation \(RAG\) System](#) | Tech Stack: Python, LangChain, SBERT

- Built an end-to-end **document ingestion pipeline** for **200+** insurance policies, generating **metadata** and **SBERT embeddings** for semantic indexing
- Engineered a high-speed **FAISS/ChromaDB vector retrieval layer** to power accurate, low-latency semantic search
- Designed a modular **RAG architecture** using **LangChain**, custom retrievers, prompt templates, and **GROQ LLM** integration
- Developed a transparent **Streamlit interface** and delivered a **scalable, cost-efficient** RAG system enabling reliable **natural-language QA** without paid APIs

[Sentiment-Based Product Recommendation System](#) | Tech Stack: Python, NLTK, Flask

- Built an end-to-end **NLP pipeline** using **tokenization**, **lemmatization**, **stopword** removal, and **TF-IDF** feature extraction on **30k+** product reviews
- Trained an **XGBoost sentiment classifier** and combined its outputs with a **user-item similarity matrix** to create a hybrid recommendation model
- Developed a **collaborative filtering engine** enhanced with **sentiment-weighted scoring** for higher-accuracy recommendations
- Deployed a **Flask inference API** and front-end for real-time, interactive recommendations, improving ranking relevance beyond rating-only systems

[Customer Segmentation Using Clustering](#) | Tech Stack: Python, Streamlit, Docker, GitHub Actions

- Engineered **RFM features** and normalized transactional data to prepare high-quality inputs for clustering
- Applied **K-Means** (Elbow + Silhouette) and **Hierarchical Clustering** with Ward linkage to generate stable customer segments, validated through **silhouette analysis**
- Built an interactive **Streamlit dashboard** for exploring clusters and exporting labeled segments
- Containerized the end-to-end workflow using **Docker**, **Pytest**, and **GitHub Actions**, delivering a **production-ready segmentation pipeline** for targeted marketing

PROFESSIONAL EXPERIENCE

Associate - Data Analyst Nov '24 - Present
Samhita Remote

- Led large-scale data engineering and automation for the ONDC program, building Python pipelines to clean, map, and consolidate datasets across **25k+ sellers & 43k+ participants**, establishing a unified source of truth
- Managed and trained a **6-member team** responsible for data validation, catalog-quality checks, and structured reporting through standardized pipelines and automated scripts
- Streamlined reporting processes by unifying monitoring dashboards & analytics layers with **Excel**, **Power BI**, & **Python**, enhancing data quality checks by **25%** & improving anomaly detection & trend-tracking workflows
- Delivered impact in Access to Finance by **automating MIS systems**, building **Power BI** dashboards, and developing **Python+SQL+Excel** pipelines for disbursement, repayment, and participant tracking
- Performed **EDA**, **anomaly detection**, **NLP-based transcription**, and feature extraction tasks to support program evaluation, problem framing, and high-signal, decision-ready insights