SONI

sonikumrai54@gmail.com | +91-8127151863 | Portfolio | LinkedIn | GitHub | Naukri

SUMMARY

Machine Learning Engineer with hands-on experience in developing and deploying ML and Deep Learning models across NLP, Computer Vision, and Generative AI applications. Skilled in data preprocessing, feature engineering, model training, evaluation, and optimization using Python, TensorFlow, PyTorch, and Scikit-learn. Worked on real-time LLM-based systems, YOLO-based CV pipelines, and retrieval-based chat systems, delivering production-ready applications using Streamlit and FastAPI. Experienced in SQL, MongoDB, AWS, and CI/CD with strong problem-solving ability, research mindset, and commitment to scalable ML systems and continuous learning.

EDUCATION

Sharda University — *Gautam Buddha Nagar, India*
Master of Science - Data Science and Analytics | CGPA: 8.978 — *August 2023 - May 2025*

MGKVP University — *Varanasi, India*
Bachelor of Science - Mathematics | Percentage: 73% — *July 2019 - August 2022*

EXPERIENCE

Python Developer Intern — *Aug 2025 – Sept 2025*
Inoday Consultancy Pvt. Ltd., Noida, India — *On-site*

Machine Learning Intern — *Apr 2025 – Jul 2025*
Team Of Keys, Noida, Gautam Buddha Nagar, India — *On-site*

ACADEMIC PROJECTS

Research-Based Learning Project — Generative Models in Machine Learning — *Sharda University*
Conducted academic research on **GANs, VAEs**, and **Autoencoders** for image generation and reconstruction. Explored unsupervised feature learning applications in computer vision as part of Master's research project.

Community Connect Project (SDG-8) — *Sharda University*
Led village surveys analyzing **100+ responses** using **SPSS** and statistical hypothesis testing. Generated actionable community development insights aligned with UN Sustainable Development Goals.

MACHINE LEARNING AND AI PROJECTS

AI-Powered Telegram Chatbot with LLM Integration
Engineered an automated chatbot using **n8n workflow automation**, **Telegram Bot API**, and **Groq's Llama 3.3 70B Versatile** model for real-time natural language processing. Implemented prompt engineering techniques and hyperparameter tuning (temperature: 0.7, max tokens: 1024) to optimize response quality and coherence. Successfully deployed production-ready system handling unlimited concurrent conversations with sub-5-second latency.

AI-Powered Dynamic Mock Interviewer — *Live App*
Developed an AI-driven mock interview platform using **Streamlit** and **Groq's Llama 3.1 70B** model. Designed a real-time evaluation system generating instant feedback across **5+ technical domains** (ML, NLP, CV, Python, Data Science). Reduced interview preparation time by **40%** through automated feedback and analytics dashboard. Deployed on **Streamlit Cloud** for scalable access.

Multimodal Emotion Recognition System
Developed real-time emotion recognition integrating facial **(DeepFace)**, speech **(SpeechRecognition)**, and text

analysis **(Transformers)**. Built interactive **Streamlit UI** with **Plotly** visualizations and implemented contradiction detection across modalities for sentiment fusion.

Sales Forecasting Chatbot — *Live App*
Built AI-powered chatbot using **Streamlit** and **Gemini API** for real-time sales forecasting. Integrated statistical models with conversational AI to deliver predictive business insights through intuitive chat interface.

Football Player Tracking System
Created real-time player detection and tracking using **YOLOv5/YOLOv8** and **Dense Inverse Search Optical Flow**. Generated tactical insights and performance heatmaps using custom **Roboflow** datasets, **OpenCV**, and **Matplotlib**.

RAG Pipeline Chatbot with Vector Database Integration
Engineered intelligent document processing system using **n8n automation**, **HuggingFace embeddings**, and **Pinecone vector database**. Achieved **75% improvement** in document accessibility through semantic search and real-time retrieval across cloud-stored files.

AI-Powered Customer Support Automation System
Deployed end-to-end customer service workflow using **n8n** and **Gmail** integration. Automated email parsing and response generation with **LLMs**, reducing response time by **90%** while handling multiple concurrent requests without human intervention.

Fine-Tuning LLaMA-2 Chat Model on Multilingual OCR Dataset
Fine-tuned **LLaMA-2-7B-chat** on multilingual manuscript data using **LoRA, QLoRA, and PEFT**. Processed OCR-extracted text with **PyMuPDF** into structured instruction format using **Hugging Face Transformers** for efficient memory-optimized training.

Autonomous Agent Development with n8n
Created multi-functional autonomous agents integrating **OpenAI, Mistral, Apollo.io**, and **Google Cloud APIs**. Automated LinkedIn content creation, invoice processing, and built customer-facing chatbots with real-time data enrichment through web scraping and RAG pipelines.

TECHNICAL STRENGTHS

Machine Learning Techniques: Regression, Classification, Clustering, CNNs, RNNs, Transformers, Fine-Tuning, Feature Engineering, Model Evaluation Metrics.

AI/ML Frameworks: TensorFlow, Keras, PyTorch, Scikit-learn, Hugging Face Transformers, DeepFace

LLMs Agentic AI: OpenAI API, Claude API, Gemini API, Groq LLaMA (GPT, Claude, Mistral), LangChain (learning), RAG Pipelines, LoRA, PEFT, LLaMA 2 Fine-tuning

Vector Databases Retrieval: Pinecone, FAISS (familiar), Chroma, Semantic Search, Document Embeddings

Computer Vision: OpenCV, YOLOv5/v8, Dense Inverse Search (DIS), Haar Cascade, CNNs

NLP & OCR: Tesseract, Google Cloud Vision API, SpeechRecognition, PyMuPDF, Text Processing

Workflow Automation: n8n (Advanced), API Orchestration, Multi-step Automation, Webhook Management

Programming & Data: Python, R, C, SPSS, Pandas, NumPy, Matplotlib, Seaborn, Statistical Analysis

Cloud & Big Data: AWS, Google Cloud Platform (Vision, Translation, TTS, STT), Apache Spark

Vector Databases & APIs: Pinecone, OpenRouter, Apollo.io, Tavily, yFinance

Web Development: HTML/CSS/JavaScript, Streamlit, FastAPI, Tkinter

Databases: SQL (MySQL/PostgreSQL), MongoDB

CERTIFICATIONS

- Data Structures and Algorithms in Python (GeeksForGeeks)- Ongoing
- Learn Python Programming - Beginner to Master (Udemy)
- Building Language Models on AWS (AWS)
- Big Data Hadoop and Spark Developer (Simplilearn)
- MongoDB Developer and Administrator (Simplilearn)

CURRENTLY LEARNING

Self-Study & Online Learning: MLOps (End-to-End Pipelines, Model Serving, CI/CD, AWS/GCP Deployment), LangChain & RAG Pipelines, Prompt Engineering for Generative AI, Experiment Tracking with MLflow