

# Sumeet Khatri

## Machine Learning Engineer | MLOps | System Design | AI-Agent

✉ sumeetkhatri1771@gmail.com

☎ +919328050693

📍 Pune

🌐 LinkedIn

🐙 GitHub

🔗 Portfolio

 Kaggle

### PROFESSIONAL SUMMARY

Machine Learning Engineer with hands-on experience in building, deploying, and scaling end-to-end AI systems. Skilled in computer vision, MLOps, and real-time model serving using FastAPI and Streamlit. Delivered AI solutions achieving 92% AUC in credit scoring and 98% accuracy in healthcare prediction, with measurable business impact.

### Projects

#### Large Language Model (LLM) Fundamentals & Construction Exploration 🔗

- Gained comprehensive understanding of Large Language Models (LLMs), covering tokenizer pipelines, dataset preparation, pretraining objectives, and model architectures.
- Explored model training workflows, optimization techniques (AdamW, mixed-precision), and evaluated trade-offs in transformer-based architectures (BERT, GPT).

#### News Summarizer, NLP 🔗

- Built an abstractive summarization tool using BartForConditionalGeneration from HuggingFace to generate concise, real-time news highlights.
- Deployed via Streamlit to provide interactive summaries of trending articles, leveraging news APIs and dynamic NLP pipelines.

#### Toxic Comment Classification Web App 🔗

- Built and deployed a multi-label NLP model to classify comments into toxic categories using BERT & HuggingFace Transformers; fine-tuned on Kaggle's Jigsaw dataset.
- Developed a real-time Streamlit interface to flag abusive language; implemented precision-optimized model validation and class balancing for production-grade moderation.

#### Credit Risk Modeling 🔗

- Engineered a credit risk scoring model using LightGBM and logistic regression, achieving 92% AUC on imbalanced datasets.
- Incorporated Weight of Evidence (WoE) transformation and hyperparameter tuning with Optuna to optimize for regulatory transparency.

#### Car Damage Detection 🔗

- Built a real-time car damage assessment system using ResNet50 for multi-class image classification (dent, scratch, broken).
- Achieved 93% test accuracy; deployed with FastAPI backend and Streamlit UI, offering sub-200ms inference time for insurance applications.

### Professional Experience

#### Software Engineer Intern-Deloitte Australia, Virtual Internship India

- Analyzed 10M+ sensor records using PySpark, improving efficiency by 40%.
- Built KPI dashboards with Power BI and automated reports
- Proposed AWS pipeline (Kinesis, Lambda, Redshift) cutting costs by 30%.
- Facilitated workshops and wireframe creation in Figma.

### Technical Skills

- **Languages:** Python, C++
- **ML & DL:** Scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM, CNNs, NLP, Transformers
- **MLOps & Deployment:** FastAPI, Streamlit, REST APIs, Git, ML Pipelines, CI/CD Pipelines, Cloud, Model Deployment
- **Data Engineering:** PySpark, Data Preprocessing, Feature Engineering
- **Foundations:** System Design, OOP, Data Structures & Algorithms (DSA)
- **Generative AI:** Feature Engineering, Development AI-Agent, Langchain, Langgraph, Vector Database, LLMs, Retrieval Augmented Generation (RAG), Model Context Protocol (MCP), CrewAI

### Education

#### B.Tech IT, Indus University

07/2022 – 07/2026 | Ahmedabad

### Certificates

#### Deloitte

Virtual internship providing hands-on experience in data processing

#### C++ DSA

Course focused on DSA concepts using C++

#### Master in Machine Learning

Completed project-based course covering core ML algorithms, pipelines, and model deployment strategies.