




SUMIT PANDEY

 [5umitpandey.github.io](https://github.com/5umitpandey)

+91 8827149288  sumitpandey@myyahoo.com  linkedin.com/in/5umitpandey  github.com/5umitpandey

Education

Thapar Institute Of Engineering & Technology

Master of Computer Applications

2024 – 2026

Patiala, Punjab

Ajeenkya DY Patil University

Bachelor of Computer Applications

2020 – 2023

Pune, Maharashtra

Experience

Amazon ML Summer of School

July 2025 – Aug 2025

Mentee

Remote

- Selected among top candidates for Amazon's prestigious Machine Learning Summer of School program.
- Gained hands-on experience in advanced ML concepts, real-world problem solving, and industry-level projects.
- Collaborated with peers and mentors from Amazon to implement scalable and optimized AI/ML solutions.

CollegeDunia

July 2023 – Jan 2024

Data Analyst Intern

Gurugram

- Automated 30+ tasks using Python, JavaScript, Selenium, and similar tools, cutting task time by over 50%.
- Extracted Insights from 200+ YouTube Analytics, Instagram Professional Dashboard and Meta Business Suite.
- Strategized & identified 20%+ growth opportunities resulting in a significant uptick in engagement and audience reach.

NCR Eduservices

Nov 2021 – Apr 2022

Tech Intern

Noida

- Successfully executed and contributed to real-life web development projects.
- Applied fundamental Data Science concepts to conduct an in-depth analysis of 10+ competitors.
- Produced 30+ captivating presentations and designed engaging content to outline the future trajectory of our website.

Projects

Secure & Intelligent Q&A System (RAG + LLM) | NLP, Vector Search | [GitHub Link](#)

Aug 2025

- Local & secure RAG system using Google Gemma, enabling private, high-performance natural language conversation.
- End-to-end data processing and retrieval workflow, text chunking, semantic embeddings, and vector similarity search.

AI-Powered Deepfake Speech Detection with Wav2Vec2 | GANs, ML, XAI | [GitHub Link](#)

April 2025

- Harnesses Wav2Vec2 embeddings and GAN-augmented training to accurately identify synthetic speech.
- Integrates XAI tools (LIME, SHAP, Grad-CAM) for transparent and interpretable predictions.

Research Paper

Text Summarisation (MultiModal) | NLP, Python, ML | [GitHub Link](#)

May 2023

- Delivers concise summaries of lengthy texts while preserving their essential content.
- Supports both English & Hindi languages, and even includes YouTube Video content.

Technical Skills

Programming Languages: Python, C/C++, SQL, Bash

Development: Data Structures and Algorithms, Software Engineering

DevOps & Cloud: Linux, Git, Docker, AWS

AI & Machine Learning: Scikit-learn, TensorFlow, PyTorch, NLP, LLM, RAG

Data Analytics: Pandas, NumPy, Seaborn, Matplotlib, PowerBI, Databricks, Spark

Certification

IBM Data Science Professional Certificate

Sept 2021

Coursera

- Tools, Methodology & Development.
- Python, Database & ML.
- Data Analysis & Visualization.