SUMIT PANDEY

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

Education

Thapar Institute Of Engineering & Technology 2024 – 2026
Master of Computer Applications *Patiala, Punjab*

Ajeenkya DY Patil University 2020 – 2023
Bachelor of Computer Applications *Pune, Maharashtra*

Experience

CollegeDunia | Data Analyst Intern **Gurugram | July 2023 – Jan 2024**
- Automated 30+ tasks using Python, JavaScript, and Selenium, reducing manual time by over 50%.
- Analyzed 200+ YouTube and Meta dashboards to uncover 20%+ growth opportunities and enhance engagement.

NCR Eduservices | Tech Intern **Noida | Nov 2021 – Apr 2022**
- Executed real-world web development projects while applying core Data Science concepts to analyze 10+ competitors.
- Created 30+ engaging presentations and content pieces outlining the website's future trajectory and growth strategy.

Projects

Social Platform for Discovery and Connections | *Elo, Graph Embeddings* | GitHub | Live Link **Ongoing**
- Dynamic social platform to foster authentic connections via real-time statuses and user-driven events.
- User-centric interface to foster meaningful connections by elevating personality and shared values over appearance.

Secure & Intelligent Q&A System (RAG + LLM) | *NLP, Vector Search* | GitHub | Test 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.

Research Paper

Text Summarisation (MultiModal) | *ML, NLP, NLTK* | GitHub | Paper Link **May 2023**
- Delivers concise summaries of lengthy texts while preserving their essential content.
- Supports English and Hindi, including YouTube video content.

Skills

Programming & Tools: Python, C++, SQL, PyTorch, TensorFlow, HuggingFace, Scikit-learn
Machine Learning & Statistics: Probabilistic Models, Regression, Classification, Clustering, Ensemble Methods
Deep Learning: NLP, Neural Networks, Transformers, LLMs, Text Classification, Embeddings, RAG
Big Data: Distributed Computing, Batch & Stream Processing, Large-scale Data Management, Hadoop
Cloud & MLOps: AWS (SageMaker, S3), GCP, Linux, Git, Docker

Virtual Internships & Certification

IBM SkillsBuild with AI | **IBM** **Nov 2025 – Present**
Vodafone Conversational Data Analysis with LLMs | **Vodafone Idea** **Sept 2025 – Oct 2025**
Amazon ML Summer of School | **Amazon** **Aug 2025 – Oct 2025**
IBM Data Science Professional Certificate | **Coursera** **Sept 2021**

Achievements

Runner-up among **250+ teams** in National Data Analytics Hackathon | **IIT Kharagpur** **Nov 2025**
6th Place among **20,000+ participants** in the Coderush Hackathon | **Indus Valley Partners** **Oct 2025**
Top 8% of **250,000+ applicants** in ATF 2025 | **AlgoUniversity** **Oct 2025**
Top 100 out of **10,000+ participants** in the Prompt Engineering Challenge | **Mercer Mettl** **Sept 2025**