Smit Thakare

(551) 344-4551 | smit.thakare@columbia.edu | https://www.linkedin.com/in/smit-thakare

EDUCATION

Columbia University New York, NY
Master of Science in Computer Science, Machine Learning Track. Aug 2025- Dec 2026

K.J. Somaiya College of Engineering Mumbai, IN
B.Tech in Information Technology (Honours in AI)**,** GPA: 9.3/10.00. Jul 2021 - May 2025

WORK EXPERIENCE

Savic Technologies Mumbai, IN
Software Engineering Intern Jan 2025 - Apr 2025

- Built NLP-powered resume parsing system using spaCy, DistilBERT, and T5-small; extracted structured insights and integrated with SAP for personalized candidate profiling, improving recruitment efficiency by 34%.
- Designed data-driven HR analytics dashboards via RESTful APIs; built scalable evaluation modules for course completions and certifications, increasing engagement by 30% through statistical analysis and insights.

Electra Enterprises Mumbai, IN
Web Developer Intern Jan 2024 - May 2024

- Developed analytics-integrated React applications; improved user engagement and traffic by 30% through data-driven performance tuning and UX optimization.
- Implemented responsive UI components and reduced page load time by 25%; enabled faster data access and improved user experience through performance analysis and optimization.

K.J. Somaiya College of Engineering Mumbai, IN
Research Assistant Jan 2023 - Jan 2024

- Built extractive summarization pipeline using TF-IDF, PageRank, and KeyBERT; achieved 85% accuracy on dense academic texts and published findings in Springer ICDMAI 2024.
- Accelerated section-wise summarization with multithreading and structured PDF parsing; reduced runtime by 14% across 50+ research papers through algorithmic optimization and scalable analysis.

PROJECTS

Dream Camera - Multimodal Photo Search System Sep 2025 - Dec 2025

- Designed hybrid retrieval framework combining CLIP embeddings (512-d vectors) and BLIP-generated captions; improved Precision@10 by 11% through 80/20 visual-textual score fusion and semantic understanding.
- Implemented multimodal clustering pipeline integrating visual similarity (70%), temporal signals (15%), and geospatial data (15%); achieved 0.38 silhouette score and 74% temporal coherence through feature engineering.

Medical Chatbot Assistant - RAG-Powered Conversational AI Feb 2025 - Apr 2025

- Built Retrieval-Augmented Generation chatbot using LangChain and Chainlit; answered health queries from 200+ unstructured medical PDFs using FAISS vector search and optimized document chunking strategies.
- Fine-tuned HuggingFace's flan-t5-base for semantic relevance; boosted answer precision by 85% and reduced latency by 40% through statistical evaluation and PyTorch optimization on Apple Silicon.

Cost-Optimized Model Routing - Complexity Prediction Sep 2025 - Dec 2025

- Trained binary classifiers on 800 mathematical reasoning problems to predict difficulty; engineered 5-dimensional complexity features achieving statistical significance ($p<0.01$) for keyword indicators through feature analysis.
- Conducted comprehensive error analysis identifying class imbalance (65/35 split) and feature separation challenges; evaluated routing accuracy (65.6%), precision (46.8%), and recall (36.4%) using confusion matrix analysis.

SKILLS

- Data Science & Modeling: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, XGBoost, LightGBM, Statistical Analysis.
- AI & NLP: PyTorch, TensorFlow, Transformers, spaCy, FAISS, Hugging Face, LangChain, Chainlit, CLIP, BLIP.
- Programming & Tools: Python, SQL, Git, FastAPI, Streamlit, PostgreSQL, Tableau, Excel, AWS.

ACTIVITIES

Joint Secretary, Emfinity Math Council Aug 2023 - Aug 2024

- Directed 27-member team hosting math events for 300+ participants; secured sponsorships and boosted attendance by 40%. Collaborated with faculty on mathematical modeling and data science workshops.