SAKET LAD

+1-(631)-366-9454 | saketlad5@gmail.com | linkedin.com/in/saket-lad | github.com/saketlad75 | Portfolio | Google Scholar

EDUCATION

Stony Brook University — **August 2025 - May 2027**
Masters of Science in Data Science — *Stony Brook, NY*
Courses: Distributed Systems, Cloud Computing, Introduction to Computer Vision, Data Structures and Algorithms

Bharatiya Vidya Bhavan's Sardar Patel Institute of Technology — **December 2021 - May 2025**
Bachelor of Technology in Electronics and Telecommunication Engineering — *Mumbai, MH, India*
Courses: Object-Oriented Programming, Database Management Systems, Computer Networks

TECHNICAL SKILLS

Programming Languages: Python, Java, MATLAB, C++, Kotlin, MySQL, NoSQL, Go, Rust, Typescript
Libraries: TensorFlow, PyTorch, Scikit-learn, Pandas, Numpy, OpenCV, Streamlit, Flask, Django
Technologies: Machine Learning, Deep Learning, Computer Vision, HuggingFace LLM's, Natural Language Processing, GenAI, Langchain, MCP, Agentic AI
Cloud & DevOps: AWS (S3, Lambda, Translate), GCP (Dialogflow), Firebase, CI/CD, Bitbucket, Kubernetes, Unix, Jenkins
Frameworks/Backend: APIs, AndroidUI, XML, JetPack Compose, RESTAPIs, PostgreSQL

EXPERIENCE

SP Jain Institute of Management and Research — **February 2025 – May 2025**
Machine Learning Intern — *Mumbai, Maharashtra*
- Enhanced **factual consistency by 34%** by fine-tuning the **BART** model on over **10-K filings** using **48 gold-standard summaries**, integrating a custom **keyword-attention mechanism** to prioritize financially salient terms
- Benchmarked **BART vs. Gemini Pro** across **32 companies**, improving summarization quality by **+8 ROUGE-L** and achieving a **BERT F1 of 81.11%**, demonstrating **human-level alignment** and outperforming baseline models by **12%**

Technocrat Solutions — **December 2024 – May 2025**
Software Engineer Intern — *Goregaon, Maharashtra*
- Optimized a production-grade **SQL Server database** for large-scale biometric data, applying **indexing**, **normalization**, and **partitioning** to improve query performance by **40%** and reduce latency under load
- Automated **Python-based ETL pipelines** for end-to-end data ingestion, transformation, and validation, reducing redundancy by **30%**, improving **data reliability**, and accelerating **retrieval speed** across analytics modules

Indian Institute of Technology, Bombay — **January 2024 – August 2024**
Machine Learning Research Intern — *Mumbai, Maharashtra*
- Engineered a mobile app for contactless fingerprint authentication using **Kotlin**, **YOLOv8x**, and **TensorFlow**, achieving **2.5% EER** and reducing latency by **20%** over **4+ deployments**
- Led the development of a **contactless fingerprint recognition** system for the Government of India, collaborating with **UIDAI** and **Mahindra Defense**, and securing $250K in funding through industry–academia partnerships
- Designed a high-accuracy deep learning model for **contactless palmprint identification**, achieving an **Equal Error Rate (EER) of 2.65%** and improving recognition accuracy by **21%** over baseline CNNs
- Implemented a **Siamese Neural Network** using **shearlets** and **contourlets** for precise feature extraction, reducing training loss and epochs by **33%** and parameters by **37%** over state-of-the-art models
- Built a custom **TensorFlow wavelet library** to incorporate **Haar and db2 wavelet backpropagation**, accelerating model convergence by **1.4x** and improving training efficiency on edge devices

PROJECTS

SafeArt: Cloud-Native Poster Compliance Platform | *Java, AWS (S3, DynamoDB, SQS, Step Functions), Docker* — **Github**
- Architected a **serverless microservice** on **AWS** using **API Gateway**, **Lambda**, **S3**, **DynamoDB**, and **SQS** for async image compliance analysis, enabling **10K+ jobs/hour**
- Added **operational resilience** idempotent requests, retry/backoff, **DLQ handling**, and minimal **CloudWatch** monitoring achieving $P95 < 2s$ cached evaluations and reducing MTTR by **60%**

ExportEase: Global Reach, Seamless Collaboration | *AWS, Python, HuggingFace* — **Demo Link**
- Engineered an end-to-end **trade communication platform** integrating **AWS Translate** and **generative AI** for multilingual summarization, reducing manual effort by **30%** and response time by **45%**
- Developed **Doc Buddy**, an **AI-powered assistant** using **HuggingFace LayoutLM** for "**Talk to PDF**", real-time Q&A, and multilingual summaries, accelerating document processing by **4x** and enhancing collaboration across **5+ countries**

Barclays ConnectCare | *Kotlin, AWS (S3, EC2), Python, GCP, NLP* — **Github**
- Innovated a **AI-driven support system** using **GCP Dialogflow** and **sentiment-aware NLP**, cutting backend latency by **35%** and improving critical issue resolution time by **45%**
- Deployed a secure **Kotlin–Firebase** android app with **AWS S3** integration, providing real-time access to **1.5K+ users** and reducing support response time by **40%**

ACHIEVEMENTS

- Selected among **60,000+** for **Amazon ML Summer School 2024** covering core machine learning concepts, optimization, and real-world use cases
- Ranked among the **top 1%** of participants at **Barclays Hack-o-Hire** coding event, where the team deployed a scalable solution for real-time financial data analysis and customer query optimization