KASHISH HETAL SHAH

+1 (857) 506-9102 | shah.kashishh@northeastern.edu | San Francisco, CA | LinkedIn | GitHub

EDUCATION

Northeastern University
Master's, Electrical and Computer Engineering
- Coursework: Fundamentals of Computer Engineering (Teaching Assistant – Fall'24), Machine Learning and Pattern Recognition, Natural Language Processing, Data Management, Small Data for ML, Parallel Processing, Data Visualization, MLOps

University of Mumbai
Bachelor's, Electronics and Telecommunication Engineering
- Coursework: Data Structure and Algorithms, Python for Data Science, Big Data Analysis, Business Analytics, Artificial Intelligence

PROFESSIONAL EXPERIENCE

KAP Ventures   **Los Angeles, CA, USA**
Machine Learning Engineer   *May 2025 - Present*
- Implemented agentic AI workflows using LLaMA and GPT to power autonomous fashion styling assistants, driving social media growth from 0 to 20K+ followers and tripling user engagement within 2 months through dynamic, conversation-led experiences
- Deployed NLP microservices via FastAPI and AWS Lambda, reducing inference latency by 40% and enhancing real-time recommendations

Northeastern University   **Portland, ME, USA**
Assistant AI Researcher   *Aug 2024 - Dec 2024*
- Researched infant neurodevelopment using 3D pose estimation models (Meta's SAPIENS, SMPL, SMIL) to support autism diagnostics in collaboration with cross-functional research teams
- Created 2D skeleton-based pipelines with Pyskl, converting 6+ hours of raw video into structured CSV datasets for behavior classification
- Processed and refined SSBD data with Pandas, NumPy, and OpenCV, reducing 50% size and accelerating preprocessing by 87%
- Supported university-led clinical research by aligning biomechanical analysis with ADOS scoring and documentation standards

Northeastern University   **Boston, MA, USA**
Data Scientist - Assistant Researcher   *Mar 2024 - Jun 2024*
- Conducted deep learning experiments using TensorFlow and Keras, increasing classification accuracy by 18% on over 500 diagnostic samples, tracked experiments with MLflow and versioned pipelines using DVC to support reproducibility and model iteration
- Organized scalable ETL flows for over 200GB sensor data using Pandas and NumPy, reducing preprocessing time by 60% and enabling deployment through Docker and Kubernetes

Kleren Oak   **Remote**
Fullstack Software Engineer   *Apr 2021 - Mar 2023*
- Launched 10+ production-ready applications using ReactJS, MongoDB and Firebase, reducing 20% delivery time and 30% load time
- Optimized backend APIs using Postman and PyTest, improving query speed by 25% and reducing bug resolution time by 40%

PROJECTS

Breast Cancer Detection Using Transfer Learning Approach - *Link to project*
- Applied CNN-based transfer-learning for cancer classification, achieving 85.48% accuracy on the MIAS data (322 images)
- Integrated VGG16, VGG19, Inception-V3, and ResNet50 with preprocessing techniques and expanded the dataset 8x through augmentation, reducing computation by 30% and lowering false positives by 15%
- Evaluated performance using F1-score (0.92), AUC, precision, and recall, contributing to AI-assisted diagnostic improvements

MindScope - Mental Health Dashboard - *Link to project*
- Built a real-time, user-centric NLP dashboard with Streamlit, processing 10K+ mental health–related tweets/day and achieving 92% sentiment classification accuracy using VADER, NRCLex, and BERTopic, while maintaining relevant documentation
- Applied critical thinking and optimization skills to architect robust preprocessing workflows with spaCy, Pandas, and NLTK, reducing execution time by 35% and enabling 6+ interactive data visualizations for social insight exploration

SKILLS

Programming Languages: Python, C++, C#, C Programming, Java, HTML, MATLAB, JavaScript, SQL, Scala, TypeScript
ML and Data Science: TensorFlow, Keras, PyTorch, scikit-learn, XGBoost, Transformers, Pandas, Numpy, OpenCV, PySpark
Data Engineering & MLOps: Spark, Hadoop, Airflow, MLflow, DVC, FastAPI, Streamlit
Web Development: ReactJS, Next.js, Node.js, Express.js, Firebase, Tailwind CSS
Databases & APIs: REST APIs, PostgreSQL, GraphQL, MongoDB
Cloud & DevOps: Docker, Kubernetes, Digital Ocean, Jenkins, CI/CD, AWS (EC2, S3, Lambda), Vercel
Visualization: Tableau, PowerBI, Matplotlib, Seaborn, Plotly