

# Vaidik Shah

Los Angeles Metropolitan Area, CA, USA

+1 (562) 682-3475 [vaidikshah03@gmail.com](mailto:vaidikshah03@gmail.com) <https://www.linkedin.com/in/vaidik-shah/> | [Github](#)

## Technical Skills

**Programming Languages:** Python, Java, C++, C#, JavaScript, TypeScript, HTML5, CSS, Go

**Backend & Databases:** Node.js, Express.js, Spring Boot, MySQL, MongoDB, Oracle, SQL Server, Cassandra

**AI/ML Tools:** NumPy, Pandas, SciPy, PyTorch (CUDA), TensorFlow, Keras, Scikit-learn, R, XGBoost, LangChain, RAG, LLM Training

**Cloud & DevOps:** AWS, Docker, Kubernetes, Git, CI/CD, Microsoft Azure, Google Cloud, Azure DevOps

**Web Development Frameworks & Libraries:** Angular, React.js, Vue.js, Bootstrap, Web3

**Tools & Visualization:** Power BI, Tableau, Matplotlib, Seaborn, Spark, Hadoop, MATLAB, Figma

## Experience

### Technolee

July 2024 – July 2025

ML Engineer

Ahmedabad, Gujarat

- Designed and deployed a **backend REST API** to serve **machine learning predictions** with **sub-200ms response latency**, enabling **real-time sales trend analysis** and **automated reporting** for an apparel company's **20+ product lines**, which **improved demand forecasting** and **reduced manual reporting time by 10+ hours per week**.
- Built and fine-tuned a **logistic regression model** on historical apparel sales data across **20+ product categories**, achieving **~80% prediction accuracy (F1 score: 0.78)** and **reducing forecasting errors by 25%** compared to baseline.
- Automated an ETL pipeline** for continuous data ingestion and **NLP-driven text preprocessing** using **PostgreSQL (SQL)**, improving **data refresh frequency** from weekly to daily and ensuring the model trained on apparel sales and customer feedback activity.

### Indian Space Research Organization (ISRO)

January 2024 – July 2024

AI/ML Engineer intern

Ahmedabad, Gujarat

- Engineered a **backend preprocessing pipeline** for **large-scale satellite datasets**, reinforcing **data governance** and ensuring **data integrity** for production model inference.
- Engineered a robust **backend preprocessing pipeline** for large-scale **satellite datasets**, reinforcing **data governance** practices, ensuring **data integrity**, and enabling reliable, high-performance **production model inference**.
- Applied **feature engineering** and **hyperparameter tuning** (cross-validation, early stopping) using **PyTorch** with **CUDA acceleration**, **reducing forecasting errors by 25%** and **improving computational efficiency by 20%**.
- Designed **clear, interactive visualizations** with **Matplotlib** and **Seaborn** to communicate **forecasting trends** and **model insights** to **scientific stakeholders**, enabling **data-driven decision-making** and **scientific reporting**.

### Cre-Art Solutions

June 2023 – August 2023

Software Engineer Intern

Ahmedabad, Gujarat

- Optimized **Django REST APIs** with **OOP principles**, streamlining financial data processing and boosting system **efficiency by 20%**.
- Implemented **data validation and optimizations**, reducing **accounting entry errors by 18%** and compliance with reporting norms.
- Built interactive Power BI dashboards** on **SQL databases** for insights into **revenue, expenses, and key metrics**, integrating **secure development practices** to ensure data integrity.

## Personal Projects

### Academic Video Summarizer / Python, FastAPI, LangChain, OpenAI (LLM + embeddings)

February 2025

- Built an **RAG pipeline** to generate context-rich summaries of academic videos by integrating external reading materials, **cutting hallucinations by ~32%**.
- Enhanced **LLM performance** through **few-shot learning** and **prompt engineering**, **improving summary by ~14%**.
- Optimized retrieval workflows using **Lang Chain's modular components**, **reducing query latency** and boosting context **relevance by ~18%**.

### Auto Review & Image Scraper / Python, Selenium, BeautifulSoup, Scikit-learn, Statsmodels

August 2024

- Built a full-stack data scraping pipeline** to collect **1,000+ real customer reviews**, **100+ product images**, and **500+ product links** from H&M's e-commerce site, creating a structured dataset for downstream ML/NLP tasks.
- Automated synthetic review generation** producing **1,200+ labelled reviews per run** with controlled sentiment distribution (75% positive, 15% negative, 10% neutral), enhancing dataset diversity and improving training data quality.

## Education

California State University, Long Beach

Aug 2025 – Present

Master of Science in Computer Science

Long Beach, CA, USA

Gujarat technological university

Sep 2020 – Jun 2024

Bachelor of Science in Computer Science

Ahmedabad, Gujarat