

Atherv Vidhate

(949)-636-9370 | athervvidhate@gmail.com | atherv.com | in/athervvidhate | github.com/athervvidhate

EDUCATION

University of California, San Diego

La Jolla, CA

Bachelor of Science in Data Science, Minor in Business Analytics

Expected Graduation: 2027

- Built Tau Kappa Epsilon alumni engagement portal using Next.js and React to support donations, networking, and access to chapter history for 1000+ members; implemented SDLC practices, conventional commits, and code review standards.

EXPERIENCE

Full-Stack Developer Intern

Aug 2025 – Present

Nexa Tax

- Delivered 95% structured data extraction accuracy in production-grade AI tax processing system with 100+ PDF documents processed through RAG pipeline, achieving structured data extraction from financial documents with 23 categories.
- Optimized computer vision pipeline achieving 60% reduction in processing time through YOLO object detection and multimodal LLM architecture for receipt/income document processing, with 100% uptime for 1000+ concurrent users.
- Engineered scalable Flask REST API with LangChain and LangGraph state management, implementing 8+ endpoints for chatbot interaction and conversation history tracking with Postgres persistence and user session management.

Software Engineer Intern

Apr 2025 – Aug 2025

smartQED

- Fine-tuned a diagnostic LLM on 200+ complex technical reports, significantly improving the model's reasoning capability and reducing root cause resolution time by 40% across product support cases.
- Integrated an AI analytics agent with remote SQLite database via Model Context Protocol, enabling real-time analysis pipelines and showcasing next-gen automation frameworks to senior leadership.
- Constructed a high-fidelity dataset for LLM training by sourcing, cleaning, and structuring over 2GB of raw aviation incident data to support high-accuracy safety prediction models.

Economics Lab Research Assistant

Apr 2025 – Aug 2025

UC San Diego Economics Department

- Engineered an open source data pipeline using the Library of Congress API to extract over 100,000 entries on historical newspaper coverage of 19th-century presidential campaigns.
- Analyzed relationships between historical newspaper titles by constructing directed graphs and applying connected components analysis to identify and group over 300 related publications.
- Developed Python scripts with API pagination and error recovery logic, ensuring 100% data retrieval completion across multi-page JSON responses and optimizing network efficiency.

PROJECTS

Fitness Program Recommendation System | <https://workout.atherv.com>

- Built a recommendation engine over 600,000+ rows with K-Means clustering, cosine similarity, and PCA for visualization.
- Fine-tuned 2 BERT models and engineered a custom reusable embedding module that outputs sentence-level vectors.
- Shipped with Docker and Google Cloud Platform for deployment and automated CI/CD pipeline with Github Actions.

Concert Scout AI | <https://concertscout.app>

- Built a full-stack AI-powered concert recommendation system using FastAPI and Next.js for over 300 active users.
- Implemented a multi-agent AI pipeline with Google Gemini to deliver personalized concert results.
- Deployed a scalable backend infrastructure with Redis session management and analytics tracking.

Multiple Linear Regression from Scratch | <https://atherv.com/linreg>

- Built a multiple linear regression model from scratch, using gradient descent to optimize the model's parameters and train/test split to evaluate model performance.
- Deployed on Streamlit to allow users to learn about MLR and input their own data and see the model's predictions.

Power Outage Analysis | <https://atherv.com/poweroutageanalysis>

- Used techniques like one-hot encoding, hyperparameter tuning, and cross-validation to build optimized Random Forest models predicting outage severity, achieving a 15% accuracy improvement.
- Cleaned 1,500+ row dataset, ensuring structured formats and engineering 4 new features for enhanced analysis.

TECHNICAL SKILLS

Programming Languages: Python, SQL, Java, TypeScript, JavaScript, HTML/CSS

Frameworks & Libraries: scikit-learn, PyTorch, Spark, LangChain, LangGraph, Dask, FastAPI, Flask, React

Tools & Platforms: Git, Docker, AWS, Google Cloud Platform, Redis, PostgreSQL, Tableau, Power BI