Chaitanya Patil

patilchait@gmail.com | +91 9270288375 | linkedin.com/in/chaitanya-patil- | github.com/chaitanyaspatil

Experience

Machine Learning Engineer, Decide Technologies Inc. – Austin, TX   Aug 2021 – Present
- Reduced model inference latency 100× by designing a JIT-compiled ad decision engine using Python and numba.
- Owned the model evaluation pipeline, unifying all evaluation logic into one reusable framework adopted across the ML stack.
- Developed a general-purpose parallel execution system using multiprocessing, achieving more than a 4× speedup in model evaluation and enabling reusable multi-process functionality across the ML stack.
- Designed a reusable reporting framework to unify metrics, automate performance summaries, and integrate with Slack and Pub/Sub for event-driven monitoring.
- Spearheaded design and deployment of four high-performing ad selection and pricing models, incorporating metric-driven filtering and blending methodologies to achieve top-quartile revenue and CTR performance.
- Worked with platform engineers to integrate models into containerized CI/CD pipelines with Docker and GitHub Actions.

Data Science Intern, Scripps Institution of Oceanography – La Jolla, CA   Jun 2020 – Jul 2021
- Migrated sensor data from Google Drive to MySQL, using indexes to accelerate preprocessing by more than 100×.
- Conducted experiments and applied Principal Component Analysis for acoustic source localization in an indoor environment.
- Generated exploratory plots using Python (Matplotlib, Pandas) to support data analysis and team investigations.
- *Publication:* 'Audio Scene Monitoring using Redundant Un-localized Microphone Arrays,' IEEE Internet of Things Journal, Volume 9, Issue 6, March 2022.

Projects

Resume RAG Chatbot (Python)
- Built a retrieval-augmented chatbot that answers questions about my background; chunked a Markdown resume, embedded with OpenAI `text-embedding-3-small`, and indexed with FAISS.
- Deployed FastAPI backend (Render) with a static UI (GitHub Pages); added CORS hardening, health checks, and token/latency logging; optimized for cost within Render's free-tier memory limits.

Time Series Caption Generator (Python)
- Built a neural captioning model using a CNN and an LSTM in PyTorch; implemented a custom noun-weighted loss function.
- Conducted systematic ablation studies; project received top grade in Neural Networks course, was praised for novelty, rigor.

Recommender System for Instacart Users (Python)
- Built a collaborative filtering system using user-product matrices to recommend items based on behavior of similar users.
- Implemented TF-IDF weighting, popularity adjustments, and product-category enrichment; visualized user trends and results.

Modified Multi-Armed Bandit using Reinforcement Learning (Python)
- Solved a complex multi-armed bandit problem involving bucket acquisition and splitting, under budget constraints.
- Empirically demonstrated an increasing profits trend over time using an epsilon-greedy reinforcement learning agent.

Skills

Languages: Python, Javascript, SQL, C

Frameworks and Tools: PyTorch, TensorFlow, Keras, Scikit-Learn, Numba, CatBoost, Pandas, NumPy, Matplotlib, Docker, Google Cloud Platform, BigQuery, Pub/Sub, Linux, GitHub, REST APIs, FastAPI, Uvicorn, LangChain, FAISS, OpenAI API, GitHub Pages, Render, GitHub Actions, CI/CD

Specializations: Recommendation Systems, Real-time Inference, JIT Compilation, Multiprocessing, Performance Optimization, Software Design, Research, Retrieval-Augmented Generation (RAG), Prompt Engineering, LLM Application Deployment

Education

University of California, San Diego, M.S. in Electrical Engineering   Sep 2019 – Jun 2021
University of Pune, B. Tech. in Electronics Engineering   Aug 2015 – May 2019

Coursework

Statistical Learning, Data Structures, Design and Analysis of Algorithms, Object-Oriented Programming, Database Management Systems, Python Programming, Deep Learning for Computer Vision, Natural Language Processing