# Aman Nagarkar

Sunnyvale, CA | 669-237-2437 | amanpnagarkar08@gmail.com | Portfolio

## EXPERIENCE

**Software Engineer - ML** Jul 2023 – Present
Frugal Hub - SCU Santa Clara, CA

- **ML Pipeline:** Developed a ML microservice healthcare application in Java, deployed it on AWS. Set up document store, data pipelines, and LLM model on Sagemaker, gaining 5000 user signups.
- **Model Evaluation:** Utilized Python for evaluating various performance metrics like pair score, precision and AUC to evalute the performance of the prediction model improving model performance by 12% and reducing type 2 errors.
- **Document store:** Utilised spaCy and, NLTK to create training corpus which can was used to create sentence embeddings using TF-IDF and Opensearch.
- **Data Pipelining:** Managed big data pipelines using PySpark to optimize data delivery, reducing data latency by 40% improving processing. Utilized and compared few-shot training vs Finetuning approach using MLLib.
- **Latency optimisation:** Utilized efficient data modeling by analysing schema and API design using RESTful principles, resulting in a 40% reduction in server response times.
- **Model deployment:** Hosted ML models as RESTful endpoints using Django and Sagemaker to be consumed by the frontend.

**Machine Learning Engineer** Jun 2022 – Sep 2022
KLA Tencor Milipitas, CA

- **ML service architecture:** Implemented end-to-end ML microservice from inception to delivery, improving global service engineer support by saving 10 hours of root cause analysis.
- **Data Ingestion:** Utilized PySpark-SQL for ETL pipeline on Snowflake to ingest data from 6 data sources to transform into a standardised format for further analysis.
- **Data Analysis:** Created a data mart for model training, conducted analysis of 8.1M data points using Python scripts and pandas, identified key feature patterns, anomalies, relationships, and trends.
- **Model selection and tuning:** Developed an automated labeling model proficient in categorizing cases according to their respective topics. Conducted A/B testing on a selection of 12 models, tuned hyperparameters for 8% enhancement in accuracy saving 200 hours of data labelling.
- **Stakeholder management:** Created interactive data visualizations using Grafana to communicate key findings and performance metrics to stakeholders, facilitating data-driven decision-making.
- **Workflow automation:** Utilized Airflow to automate workflow for organising and storing the vector embeddings from input queries, reducing data processing time by 30%.

**Software Engineer - ML** Feb 2019 – Jul 2021
Vint Media Pune, India

- **AOV Improvement:** Leveraged transaction data to design and deploy a content based product recommendation system leading to a 20% improvement in AOV.
- **Data pipeline optimisation:** Managed customer data pipeline with EMR and Spark to optimize data delivery for efficiency to process 550GB of data. Used Kafka for faster retrieval by 10%.
- **Churn prediction:** Trained a XGBoost classifier using sklearn for churn prediction, achieving 82% precision.
- **Query Optimisation:** Analysed and optimised SQL queries using CTE's to reduce redundant database calls improving loading times by 15%.
- **Service deployment:** Collaborated in a crossfunctional enviorment to containerize and orchestrate machine learning microservices in production using **Docker and K8s**
- **Software Testing:** Performed code reviews, wrote comprehensive testcases and used JIRA for bug tracking, issue tracking and project management.

## SKILLS

**Languages & Databases**: Python, Java, Scala, Golang, JavaScript, SQL, MongoDB, PostgreSQL, GraphQL.
**Deep Learning Frameworks**: PyTorch, TensorFlow, HuggingFace, Scikit-learn, XGBoost, Spark, Kafka, PySpark, Snowflake.
**Technologies**: AWS (SageMaker, EC2, DynamoDB), Opensearch, Spark, Kafka, Jira, Grafana, Kubernetes, Hadoop, Airflow, MLLib, Git, CUDA.

## EDUCATION

**M.Sc - Computer Science and Engineering** Sep 2021 – Jul 2023
Santa Clara University - Santa Clara, CA

**B.E - Computer Science** Aug 2015 – Jul 2019
Savitribai Phule Pune University - Pune, India

## PROJECTS

**Multimodal RAG using Claude-3** | AnthropicMultiModal, LlamaIndex, Python

- Utilized Llama Index to act as an orchestrator to perform multimodal tasks using Claude-3 by Anthropic. Used Uber 10Q dataset to parse over a pdf for text retrival using LLamaParser. Extracted out text from images using the parser. Used Query engine to create a Multimodal RAG system to query on new data.

**Document Query Engine using Llama 2.0** | FastAPI, Cuda, Pytorch, Llama2, Langchain, ChromaDB

- Created a ML pipeline leveraging the RAG (Retrieval-Augmented Generation) architecture on Llama 2.0 LLM as the generator component. Orchestrated both the retriever and generator elements using Langchain.
- Used a local instance of ChromaDB to serve as a vector database, for efficient data retrieval and storage within the system. Designed RAG pipeline on Kubernetes for real-time query handling.