NAMAN DALSANIA

github.com/namandalsania | linkedin.com/in/naman-dalsania | namandalsania12@gmail.com | +1 213-681-4183

EDUCATION

University of Southern California Jan 2024 - Dec 2025
Masters in Applied Data Science Los Angeles, CA
Courses: Machine Learning, Data Mining, Predictive Analytics, Data Management, Research Methods & Analytics for User Studies, Fairness in AI, Natural Language Processing

TECHNICAL SKILLS

Programming language: Python, SQL, SAS, MATLAB, Java, JavaScript
ML/AI: TensorFlow, PyTorch, OpenCV, Apache Spark, Apache Hadoop, MapReduce
Web Technologies: React.js, Node.js, MongoDB, Django, Flask
Tools: AWS, Tableau, LINUX, Power BI, Anaconda, SPSS, Excel, Git, Bash

EXPERIENCE & RESEARCH

Data Science Intern |MLOps (MLflow), Data Orchestration (Databricks, ADF), Cloud Infrastructure (Azure ADLS) Aug 2025 - Dec 2025
Kaiser Permanente Pasadena, CA

- **Orchestrated MLOps pipelines** using **Azure Data Factory** and **Databricks**, automating the ingestion and transformation of skewed **EHR** data while establishing a standalone baseline using log-transformed targets and percentile binning.
- **Engineered a Mixture-of-Experts framework** within a **CI/CD-driven MLflow lifecycle**, implementing hybrid Gamma/Huber loss and isotonic calibration to isolate outliers and achieve a 15% weighted RMSE reduction in production-ready models.

Machine Learning Engineer Intern | Systems Programming (C++, Rust), Cloud-Native AI (GCP, K8s), LangChain May 2025 - Aug 2025
DataDirect Networks San Francisco, CA

- Developed and optimized a **C++-based NVIDIA's NIXL** (NVIDIA Inference Xfer Library) plugin enabling RDMA-based GPU memory transfers with Infinia **KV store**, accelerating inference for **vLLM**, **TensorRT**, and **SgLang** on A100 and L4 GPUs.
- Designed **async data flow logic**, benchmarked **LMCache-based LLM serving** on **Google Cloud Platform (GCP)**, and instrumented detailed traces using Perfetto, REDview, Heaptrack, and Hotspot to identify system-level bottlenecks.
- Built a **Retrieval-Augmented Generation (RAG)** pipeline supporting multi-file PDF ingestion, re-ranking, and LLM interaction using **LangChain, FAISS,** and **NVIDIA NIM (NVIDIA Inference Microservices)**, with a shared knowledge base and streaming responses.

Research Assistant | Tableau, Excel, Streamlit, Plotly Jan 2025 - Dec 2025
USC Schaeffer Center for Health Policy & Economics Los Angeles, CA

- Automated the processing of 15+ years of Medicare claims data to create **interactive Tableau dashboards**, accelerating policy reporting and supporting **National Institute of Aging (NIA)** presentations on **geographic and demographic dementia cost disparities**.
- Built a data-driven dashboard using Python, Streamlit, and Plotly to **simulate public health policy interventions**, **visualizing the potential reduction in dementia and chronic diseases** across the US population by **integrating complex regression models** into an intuitive user interface.

Research Assistant | LLMs (BERT), PyTorch, CUDA, GPU Optimization May 2024 - Dec 2024
USC Marshall School of Business Los Angeles, CA

- **Architected a low-latency NLP pipeline** leveraging **Hugging Face Transformers** and a fine-tuned **BERT** architecture to transform unstructured voice inputs into structured financial ledgers with high-fidelity classification..
- **Engineered high-performance inference optimizations** on **HPC nodes**, implementing **CUDA-based GPU parallelization** to accelerate processing speeds by 60% and reduce inference latency by 40% for large-scale production workloads.

Data Science Intern | OpenCV, OCR, API, Tableau May 2024 - Aug 2024
FairView Capital West Hartford, CT

- Architected a **Computer Vision-driven OCR solution** that streamlined financial reporting for investee funds, slashing manual overhead by **80%** and significantly enhancing processing precision for high-stakes investment data.
- **Streamlined cross-platform data synchronization** by developing automated workflows via **PitchBook's REST API**, reducing analyst data-entry bottlenecks by **72%** and ensuring real-time database accuracy within DealCloud.

PROJECTS

YelpBoost: Predictive Engine for User Ratings | PySpark, Recommender Systems Sep 2024 - Nov 2024

- Built a **hybrid recommendation system** using **XGBoost** and **CatBoost** on top of **PySpark RDD**s to predict user-item ratings in the Yelp dataset.
- Engineered **behavior-based features** and applied **dimensionality reduction** techniques to improve prediction accuracy by **12%**, ensuring full pipeline scalability using distributed Spark operations.

Credit Card Approval Prediction using Classification Algorithms | Python, Streamlit, Scikit-learn Jun 2022 - Nov 2022

- Built a machine learning pipeline to evaluate credit card application decisions, using a Gradient Boosting Classifier with **84.2%** accuracy and **0.90** Precision/Recall scores.
- Implemented a meta-learning framework across 10 predictive models and derived actionable business insights through **univariate** and **bivariate** analysis.

POSITIONS OF RESPONSIBILITY

Business Team Associate, USC Racing May 2024 - May 2025

- Led USC Racing's sponsorship, finance, and marketing efforts, boosting revenue and strengthening brand presence.

Venture Scout, Pegasus Angel Accelerator Mar 2024 - May 2024

- Worked with investors to assess startup pitches, offering strategic insights and strengthening due diligence.