NAMAN MEHTA

mehtanaman1199@gmail.com / linkedin|naman101199 / naman101199.github.io/portfolio

EDUCATION

The University of Chicago Chicago, IL
Master of Science in Applied Data Science – 4/4 GPA (Top 10% of class)

Narsee Monjee Institute of Management Studies Mumbai, India
Bachelor of Science in Applied Statistics & Analytics – 3.58/4 GPA

TECHNICAL SKILLS & CERTIFICATIONS

- **Programming Languages & Databases:** Python, R, SQL, NoSQL (Neo4j, MongoDB), C++, HQL
- **Cloud Tech Stack:** AWS, Microsoft Azure, Databricks
- **Frameworks & Libraries:** Pandas, Scikit-learn, PyTorch, Spark, Kafka, Streamlit, Flask, LangChain, TensorFlow
- **Platform Tools**: GitHub, Airflow, Docker, Kubernetes, Rest APIs, Shell Scripting, Tableau, PowerBI
- **Certifications:** AWS Certified Machine Learning – Specialty

WORK EXPERIENCE

Jones Lang LaSalle | Senior Business Intelligence Engineer United States | Jan '24 - Present

- Responsible for implementing predictive maintenance models and designing advanced analytics strategies on 10 TB of Energy Storage Systems data, leveraging Azure Data Lake and PySpark in Azure Databricks, reducing system downtime by 35%.
- Developed and deployed machine learning models using Azure Databricks, and Azure ML Studio, which optimized facility management operations. Achieved a 40% reduction in high-priority work orders, and a 15% reduction in budget usage.
- Implemented a sentiment analysis system for 10,000+ user feedback for QR codes in 1000+ buildings using Python and Azure Text Analytics, improving user satisfaction by 20% and response time to issues by 30%.

Jones Lang LaSalle | Data Engineer Intern United States | Jun '23 – Dec '23

- Designed and engineered a real-time alert notification system and weekly summary dashboard for 200 GB of IoT data from 400+ office and retail locations, integrating into Azure Data Lake, resulting in a 50% reduction in incident response time.
- Developed and optimized ETL pipelines using Apache Kafka and Apache Spark in Azure Databricks, facilitating efficient data storage from bill process streams within a medallion architecture, improving data processing speed by 30%.

Pangea Tech | Data Analyst India | Nov '21 - May '22

- Consolidated monthly financial planning reports for SaaS companies by developing ETL process using AWS Lambda, AWS Glue ETL, and AWS RDS to spotlight trends and train predictive ML models in AWS SageMaker for customer churn and root cause analysis, resulting in annual savings of over $200K.
- Integrated a sales forecasting model using Amazon S3 for dataset management, Athena for efficient SQL querying, and SageMaker for accurate time-series analysis, culminating in actionable insights through Tableau dashboards for vaccine sales strategy.
- Enhanced Tableau dashboards with row-level security, enabling granular data access control, which resulted in a 30% increase in data confidentiality compliance across various organizational levels.

Equity Data Science | Quant Data Analyst India | Jun '20 - Oct '21

- Delivered customized solutions to track investments, risk, and return through ETL pipelines and various Spotfire dashboards to analyze performance, resulting in a 90% POC to client conversion and 100% subscription from existing clients.
- Architected an automated email reporting system for clients to consolidate their portfolio and market data. Generating additional revenue from 70% of existing clients.
- Solved clients need to get insights from large financial documents by implementing an NLP-based analytical system which added $200K additional revenue.
- Managed the creation of a detailed data dictionary, which facilitated improved data modeling and patient tracking analytics.

PROJECT WORK

Generative AI Researcher (Text Query Language)

- Engineered a GenAI product that generates SQL query from natural language by fine-tuning Large Language Models (LLMs) on GCP leveraging Vertex AI Studio, Model Garden, Hugging Face, and QLoRA. This reduced manual data querying time by 80%.

Forecasting NYC Yellow Taxi Prices and Demand

- Conducted detailed exploratory data analysis (EDA) on around 29 GB of NYC Taxi trip data via Spark SQL on GCP; uncovered key demand trends and price sensitivities, leading to strategic fare adjustments and revenue growth.
- Formulated highly accurate taxi demand forecasts by training Random Forest and XGBoost models on trip activity data using Spark ML; optimized fleet allocation, reducing idle time by 18%.

Building Algorithmic Trading Infrastructure

- Developed an end-to-end algorithmic trading bot using Google Cloud Platform (GCP), leveraging VMs, BigQuery, Cloud Functions, and Pub/Sub to process real-time market data, execute trades, and manage risk, achieving a 15% improvement in trade execution speed.