MONISHVER CHANDRASEKARAN

+1 551-392-5962 | monishverchandrasekaran@gmail.com | linkedin.com/in/monishver | monishver11.github.io | Data + ML Blogs

EDUCATION

New York University, Courant Institute of Mathematical Sciences — New York City, NY
Master of Science in Computer Science, AI Concentration. GPA: 3.72/4 — *Sep. 2024 – May 2026*
Recitation Leader: CALC I, MATH-UA.121.019 & MATH-UA.121.020 (Fall 2025)

Amrita School of Engineering, Amrita Vishwa Vidyapeetham — Coimbatore, IN
Bachelor of Technology in Computer Science — *Aug. 2018 – June 2022*
Cumulative GPA: 9.37/10.00, First class with Distinction

WORK EXPERIENCE

Data Engineer — August 2022 – August 2024
IBM - Chief Information Office (CIO) — *Bangalore, IN*

- Led data engineering initiatives as Product Owner for the GHD business unit, optimizing and scaling data pipelines to drive predictive analytics and data mining capabilities.
- Migrated IBM's transactional data warehouse to IBM Cloud using Apache Spark and Scala, building ETL for batch ingestion.
- Completed Fact & Dimensional modeling and migration of the legacy data warehouse to a modern, fault-tolerant system, reducing costs by **81.3%** and decommissioning the legacy system within a strict deadline.
- Optimized and reduced SQL scripts from **600+ to 22 ETLs**, simplifying job execution through new modeling and architecture.
- Implemented data lakes and data marts for structured data storage, improving processing speed by **63.6%** and establishing a robust infrastructure for advanced business analytics and strategic insights.
- Developed Jenkins CI/CD pipelines and Apache Airflow DAGs for daily/weekly deployments, automating 15 deployments per month, with customized logging for enhanced monitoring and alerting, reducing downtime by **37.5%**.
- Participated in a PoC for Change Data Capture (CDC) using Apache Kafka and Debezium, transitioning from batch to streaming data pipelines for improved real-time data flow.
- Created an internal automation tool to ease the migration to IBM's unified ingestion framework, reducing manual workload.

Software Developer Intern — January 2022 – July 2022
IBM - Chief Information Office (CIO) — *Bangalore, IN*

- Assisted with 6+ cross-functional teams to streamline data workflows, integrating data pipelines that supported efficient ML algorithm development for optimal bid pricing.
- Validated and refined 40+ SQL scripts into 6 Fact models, ensuring high-quality, structured data for accurate reporting.
- Documented 17 key business logic for bid pricing modules by analyzing existing SQL scripts and archived tables in DB2.
- Automated Datamart schema generation with a Python script, reducing DBA effort and time significantly.

Research Assistant - AMUDA Lab — June 2020 – January 2021
Amrita School of Engineering — *Coimbatore, IN*

- Performed EDA on real-time indoor localization data using BLE beacons, applying ML techniques to derive actionable insights.

PROJECTS

Gaze-Guided Reinforcement Learning for Visual Search | *PyTorch, Stable-Baselines3* | **Project Link** — April 2025 – May 2025

- Implemented a novel RL framework integrating human gaze patterns with PPO algorithm in AI2-THOR using three integration methods (channel, bottleneck, weighted) and custom CNN architectures to process 4-channel inputs (RGB + gaze heatmaps).
- Achieved 26% better performance than random baselines across multiple floor plans with improved sample efficiency in 3D visual search and object detection tasks.

MTA Transit Time Prediction | *Pandas, NumPy, Sklearn, Seaborn, TensorFlow* | **Project Link** — Oct 2024 – Dec 2024

- Designed robust regression models to predict NYC bus travel times using MTA BusTime and TomTom Traffic data.
- Devised a grid-based modeling approach with XGBoost that captured localized traffic patterns, achieving MAE of **43.73** seconds.
- Evaluated LSTM architectures to optimize short-sequence temporal data predictions, attaining an RMSE of 132.23.

Health Insurance Cross-Sell Prediction Case Study | *Pandas, NumPy, Scikit-learn* — March 2021 – May 2021

- Engineered a predictive machine learning model to forecast customer propensity for purchasing additional insurance products, using Random Forest to increase Recall score by 27% to 0.9063 and achieving the highest F1 score among all tested models.

TECHNICAL SKILLS

Programming: Python, C++, Scala, SQL (DB2)
Frameworks: Apache Spark, Apache Airflow, TensorFlow, Keras, PyTorch, OpenAI Gym, FastAPI, dbt (Data Build Tool)
Cloud: AWS (EC2, S3), GCP
Developer Tools: Git, Docker, Kubernetes, Bazel, Jenkins, LogDNA, VS Code, Jupyter Notebook, Anaconda
Libraries: Pandas, NumPy, Matplotlib, Scikit-learn, XGBoost, SpaCy