

EDUCATION

University of Pennsylvania (School of Engineering and Applied Science)

Philadelphia, PA

Master of Science in Engineering (MSE) in Data Science; GPA 3.9/4.0

Aug 2023 – May 2025

Coursework: Database & Information Systems, Adv. Deep Learning (Diffusion models, LLMs, Multimodal architectures), Generative Modeling, Adv. Machine Perception, Machine Learning, Statistical/Probability models in Marketing

Kirori Mal College (KMC – University of Delhi)

Delhi, India

Bachelor of Science (BS) in Statistics; GPA: 8.11/10

July 2017 – Aug 2020

EXPERIENCE

Research Intern, [Computational Social Listening Lab \(UPenn\)](#)

May 2024 – Present

Misinformation: Working on identifying misinformation on social media and whether it relates to health outcomes.

- Extracted linguistic features from posts using [DLATK](#) and applied LDA for topic modeling and performed correlation analysis.
- Conducted entailment analysis using a pre-trained **RoBERTa** model to detect misinformation by post alignment with trusted claims.
- Unified data from various online survey platforms in a secured server via MySQL and performed feature engineering in Pandas to prepare data for further downstream tasks.

IH Risk Model (ongoing): Working on developing a model to predict the risk of Incisional Hernia (IH) in patients' post-surgery by using real-world operative notes and intraoperative EHR data.

- Engineered an **NLP pipeline** that integrates **GPT embeddings** with surgical data and optimizing model fitting using AutoML.

Data Science Intern, [Universal Media \(PA, USA\)](#)

May 2024 – Aug 2024

- Architected Azure SQL Database solutions, encompassing DDL scripts to enhance data management and reporting solutions.
- Led the development of **3+ data pipelines** using Azure Data Factory (ADF), facilitating the seamless ingestion and transformation of diverse data sources into the Azure environment.
- Developed python scripts for data transformation, stored them in Blob storage and executed them via batch activity in ADF.
- Used Azure Synapse to perform **Mixed Media Modeling (MMM)**, analyzing marketing channel impacts on media diversity metrics. Built **Power BI** dashboards to deliver actionable insights for optimizing client strategies.
- Authored **5 stored procedures** in SQL, automating repetitive tasks and improving query performance by over 30%.

Assistant Manager, [IIFL Finance Ltd](#)

Apr 2022 – July 2023

- Analyzed ETL process failures by familiarizing myself with Azure Data Factory and created **10+** paginated reports leveraging SQL Server Report Services functionalities to help the senior management track the business performance of 1000+ branches.
- Optimized & migrated complex SQL queries from an obsolete database server that improved the **reporting services by ~40%**.
- **Digital Adoption:** Performed analysis to ascertain the features of the customers opting for digital loan disbursements. Built a Random Forest model on Azure ML to scope out potential customers and drive digital revenue with an **accuracy of 90%**.

SELECTED PROJECTS

- **Diffusion Transformer (2025):** Implemented PatchVAE with convolutional encoders and patch-based decoding for fine-grained feature extraction. Trained a Diffusion Transformer to sample from the latent space of PatchVAE, achieving a 30% reduction in FID score compared to VAE-generated samples, demonstrating superior diversity and image quality. [\[Github\]](#)
- **Azure ETL (2024):** Built a scalable ETL pipeline to process and analyze data efficiently. Ingested raw data into Azure Data Lake Storage Gen2 using Azure Data Factory, transformed it in Azure Databricks with cleansing and feature engineering, and analyzed it in Azure Synapse Analytics, delivering a seamless end-to-end solution. [\[Github Link\]](#)
- **FitBit(2024):** Engineered a Django health chatbot leveraging PostgreSQL for robust patient data management, featuring an LLM-agnostic architecture with seamless model switching via Langchain. Optimized memory usage for handling long conversations efficiently and implemented advanced entity extraction to dynamically enhance medical context and automate request escalation. [\[Github Link\]](#)
- **Statistical Segmentation (2024):** Fitted statistical models like the Pareto II and Weibull segmented models for segmenting the customer base of movie-goers based on the unit sales of a movie dataset. Used MS Excel Solver to run the models for optimum parameters leading to future forecasts. (Solver, Marketing). [\[Link\]](#)

TECHNICAL SKILLS

Programming Languages: Python, C/C++, MySQL, R programming, JavaScript

Databases/Frameworks: MySQL, PostgreSQL, SSMS, MongoDB, Neo4j, PyTorch, React, NodeJS, HTML/CSS, Neo4J, PySpark

Cloud/Big Data Orchestration: AWS (S3, Glue), Azure (Data Factory, Synapse, DevOps), Docker, DataBricks, Kafka, Airflow, DBT

EXTRA-CURRICULAR and ACHIEVEMENTS

Theatre: Core Team Member (The Players). Lead Actor and auditioned for a brand advertisement and two short films.

Community: Collaborated with National Service Scheme (NSS), Delhi University & worked toward child empowerment.