Keyush Shah
+1 (267) 278-1269| keyush.shah12@gmail.com | LinkedIn | Website | Blog

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

Kirori Mal College (KMC – University of Delhi)   **Delhi, India**
Bachelor of Science (BS) in Statistics; **GPA: 8.11/10**   **July 2017 – Aug 2020**
<u>**Coursework**</u>: *Adv. Probability, Calculus, Linear algebra, Real Analysis, Statistical Inference, Time Series, Op. Research*

EXPERIENCE

Research Intern, ***Computational Social Listening Lab (UPenn)***   ***May 2024 – Present***
- <u>Research Goal:</u> Working on identifying misinformation on social media (FB posts, YouTube interactions) and whether or not it relates to health outcomes for people categorized on the basis of race (white/black) & urbanity (rural/urban).
- Developed a URL based misinformation-detection model that detected around 1k misinformed posts out of 1mm FB posts.
- Implemented sentiment analysis using a pre-trained '**DistilBERT'** Transformer model to evaluate user emotions from social media posts, leading to insights on engagement patterns segmented by race and urbanity.
- 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.

Data Science Intern, ***Universal Media (PA, USA)***   ***May 2024 – Aug 2024***
- Architected and implemented Azure SQL Database solutions, encompassing data pipelines, ETL processes, and DDL scripts to enhance data management and reporting capabilities.
- Led the development of **3+ data pipelines** using Azure Data Factory, facilitating the seamless ingestion and transformation of diverse data sources into the Azure environment.
- Authored **5 stored procedures** in SQL, automating repetitive tasks and improving query performance by over 30%.
- Developed and maintained 2 Power BI dashboards, utilizing DAX and Power Query to optimize data models, which provided actionable insights to client strategy teams.

Assistant Manager, ***IIFL Finance Ltd***   ***Apr 2022 – July 2023***
- Created **10+** paginated reports on Power BI Report builder leveraging SSRS (SQL Server Report Services) functionalities to help the senior management track the business performance of 1000+ branches across the country.
- **Key Projects:**
- **Digital Adoption:** Performed analysis to ascertain the features of the customers opting for digital loan disbursals. Built a Random Forest model on Azure ML to scope out potential customers and drive digital revenue with an **accuracy of 90%.**
- **Migration:** Optimized & migrated complex SQL queries from an obsolete database server to a new one with zero downtime or loss of data that improved the **reporting services by nearly 40%.**

PUBLICATIONS (PREPRINTS)

[1] Enhancing Retrieval in QA Systems with Derived Feature Association (**Under Review** at IEEE Cloud Summit)
Keyush Shah, Abhishek Goyal*, Isaac Wasserman* [ArXiv] [Github]

SELECTED PROJECTS

- **Marvel Bot (2024):** Engineered a conversational LLM app using LangChain and RAG, integrating Ollama to optimize three quantized models. Web scraped 4+ Marvel Wikipedia pages and deployed the solution on Streamlit, ensuring real-time, context-aware responses and efficient memory usage. [Github Link]
- **Traversability Estimation (2023):** Developed a terrain classification model using Semantic Segmentation and an attention-enhanced Fully Convolutional Network, achieving a 2% improvement in IoU. Enhanced off-road navigation for autonomous vehicles by optimizing path planning and terrain adaptability. (Pytorch, SkLearn, Google Colab) [Github Link]

TECHNICAL SKILLS

Programming & Skills: Python, MySQL, Java, R, Git, React, NodeJS, MongoDB, MS office, MATLAB, JavaScript, HTML/CSS
Software & Tools: PyTorch, Tensorflow, sklearn, scikit-image, OpenCV**,** VS code, Eclipse, Cloudera, Jupyter, AWS Sagemaker, AWS Bedrock, DAX Studio, SSMS, CUDA, LaTeX

EXTRA-CURRICULAR and ACHIEVEMENTS

Theatre: Core Team Member (The Players). Lead Actor and auditioned for a brand advertisement and two short films.
Sports: District-level Cricket player & represented my college in Basketball at **the zonal level**.
Community: Collaborated with National Service Scheme (NSS), Delhi University & worked toward child empowerment.