

YASH MALVIYA

774.232.5158 | ymalviya@wpi.edu | Greater Boston, MA | [LinkedIn](#) | [Portfolio](#)

VALUE PROPOSITION

- “I help organizations drive revenue and optimize user experiences through advanced ML models and data-driven strategies”.

LinkedIn Top Data Science Voice & Data Scientist with 3+ years of experience in developing and implementing AI Techniques to solve problems across healthcare, retail, and financial sectors.

SKILLS

- **Languages/Packages:** C, C++ , Data Structures and Algorithms , Python, SQL, Scikit-learn, , pyTorch, PySpark, BERT, Transformers, Apache Hive, TensorFlow
- **Tools:** Deep Learning, Machine Learning, Excel, Tableau, Git, AWS Cloud services, data bricks, Snowflake, Open Source LLM.
- **Applications:** Financial Analytics, Healthcare Analytics, Data Visualization, Generative-AI research, ChatGPT Turbo API, Docker, SLURM, MLops.
- **Soft Skills:** Stewardship mindset, intensely curious, Effective communications, People Management.

EXPERIENCE

DATA SCIENCE INTERN

May 2024 – Aug 2024

[Avality Clinical Solutions](#), Jacksonville, FL

- Spearheaded in-depth data analysis across 25+ healthcare databases, applying advanced analytics techniques to uncover trends and identify healthcare data dirtiness.
- Developed and deployed deep learning models, including GAN and Diffusion models, to create synthetic databases on AWS Terraform, enhancing data-driven decision-making.
- Collaborated with 10+ data scientists and DevOps professionals to deploy GAN models on AWS, improving synthetic data quality by 18% and generating \$1.2 million in revenue.

RESEARCH ASSISTANT

Dec 2023 – May 2024

[Worcester polytechnic institute](#) and [Avality Clinical Solutions](#), Worcester, MA

- Researched over 40+research papers and journals in 5 months to identify the gaps in current synthetic data generation techniques.
- Achieved a 12% data quality improvement, as measured by synthetic data accuracy, by developing a Diffusion model with a binary matrix conversion step, resulting in 99.89% clean data.

DATA SCIENTIST

Jul 2021 – Aug 2023

[Tata Consultancy services](#), India

- Identified a significant 15% decline in year-over-year business sell-through rates, which was negatively affecting profitability and reducing sales performance across multiple regions.
- Solved this problem by applying the Stochastic Gradient Descent algorithm to an AI model, improving forecast accuracy by 22% and enhancing sales strategy decision-making processes.

DATA VISUALIZATION INTERN

Feb 2021– Jun 2021

[Xoriant Solutions](#), India

- Detected inefficiencies in data visualization and metrics tracking that hindered data analysis processes by 40%.
- Addressed this issue by deploying ELK Stack Dashboards to provide real-time data visualization and metrics tracking.

EDUCATION

Master of Science in Data science | Worcester polytechnic Institute, MA May 2025(Anticipated)
Graduate student and Research Assistant, member, Denise Nicoletti Award Committee

Bachelor of Technology in Computer science | RGPV, India May 2017-Aug 2021
Major in Machine Learning & Deep Learning.

ACADEMIC PROJECTS

Risk Analysis & ROI Prediction on MAANG Stocks– WPI May 2024 – Aug 2024
Description: Risk and rate of Interest prediction based on stocks data with monte-Carlo simulation methods.

New York Times Headline Generation- WPI Dec 2023 – Apr 2024
Description: Using RNN and Transformer models that leverage sequence generation and Identifying clickbait Headlines by Sentiment analysis and generating Headlines for the news using their news points.

Profitability Analysis CVS-Health – WPI
Description: Performed Profitability Analysis for Max-Fashion, leading retail chain of Dubai UAE, Utilizing interactive dashboards of Tableau , Identifying KPI and Metrics that drives sales , and Tableau prep for Data preparation.

Medical Imaging -[US National Institute of Health](#) Aug 2023 – Dec 2024
Description: With 35000 Annotated Images by radiologists , created Algorithms to classify Pneumonia , Lung tumor Segmentation , Using Core Deep Learning using Pytorch and Tensorflow which Leverages cuda , cudnn

PUBLICATIONS & CONTENT CREATION

- LinkedIn Articles: Published articles in Data science community on LinkedIn in reaching over 2000+ impressions.
- Voice-Controlled Home Automation and Real-Time Object Tracking: Published in IJSREM, June 2021 (ISSN: 2582-3930) [Paper Link](#)

INTERESTS

Currently learning Spanish language - Volunteering at Bio Life Plasma Donation Camps