

VARSHIL SHAH

Computer Science Grad & AI Enthusiast

varshilshah0203@gmail.com | [LinkedIn](#) | 602-517-2465 | github.com/varshil1 | varshil1.github.io

EDUCATION

Arizona State University

Master of Computer Science, MCS (GPA: 3.8/4.00)

Tempe, AZ

Expected Graduation: May 2025

Ahmedabad University

Bachelor of Technology in Information and Communication Technology, ICT, (GPA: 3.21/4.00)

Ahmedabad, India

May 2022

WORK EXPERIENCE

TD SYNnex

Data engineering and Analytics Intern

Gilbert, AZ

May 2024 – Present

- Developed and optimized a Retrieval-Augmented Generation (RAG) LLM Bot using OpenAI GPT-4 with Databricks and Cortex, enhancing data retrieval efficiency through the implementation of FAISS algorithms and semantic caching.
- Built and maintained a robust data pipeline for preprocessing and analyzing large datasets in Snowflake, resulting in optimized performance and reduced costs for LLM calls, demonstrating strong analytical and problem-solving skills.
- Created an anomaly detection model in Snowflake, which monitors data table freshness and triggers real-time alerts via Microsoft Teams if anomalies are detected, ensuring timely resolutions to prevent business impact.
- Conducted comprehensive cost and performance analysis of different architectures, identifying Cortex as the optimal solution for cost-efficiency and performance, reflecting critical thinking and decision-making skills.

Center on Technology, Data, and Society (CTDS), Arizona State University

Research Associate

Tempe, AZ

October 2023 – May 2024

- Integrated Planet imagery seamlessly with establishment-level datasets, building pipelines to feed geospatial data to AI models for land conversion recognition and finding detailed insights into global shocks' effects on establishment dynamics.
- Employed cutting-edge technologies like SAM (Segment Anything Model) by meta-AI and GEE map in Python to analyze Planet imagery, enhancing efficiency by 15%. Contributed to the understanding of urban development dynamics and their impact on the local economy by analyzing industry relocation patterns, resulting in actionable insights for decision-makers and researchers.

Anicca Data Science Solutions

Data Engineer

Bellevue, WA

August 2022 - July 2023

- Orchestrated design and execution of high-performance data pipelines using AWS Glue and PySpark, processing datasets of 100K to Millions of rows. Implemented ETL processes, data quality checks, monitoring for relational databases and dashboards accuracy.
- Collaborated with cross-functional teams in an Agile environment to understand business requirements and create efficient Big data analytics solutions employing AWS EC2, S3, Airflow DAGs, Redshift, reducing computational time and saving \$20,000 monthly.
- Elevated data engineering operations, achieving 70% reduction in computation time. Data exploration, profiling, mining, modelling, partitioning distributed storage strategies enhanced Data Lake performance safeguarding integrity and data governance.
- Engineered a successful infrastructure migration from on-premises to cloud-based, leveraging Distributed Systems, for achieving low cost, amplifying scalability. Pioneered planning of AWS-based CI/CD system, streamlining production and testing processes.

RESEARCH EXPERIENCE

Performance Comparison of Deep Neural Networks for classification of Indian Coins

IEEE Xplore

- Designed and executed a machine vision model to precisely classify Indian coin denominations from cellphone camera images, achieving a 97% accuracy and streamlining coin segmentation work processes.
- Compared 10+ deep neural CNN network architectures including ResNet, AlexNet, GoogleNet, MobileNet, Inception, and VGG16 along with quantitative data analysis and optimization algorithms for ranking optimal classification model.

SOFTWARE PROJECTS

Social distancing detector using OpenCV

Python, Deep Learning, Computer Vision, Flask

- Developed computer vision software with prediction capabilities to detect social distancing violations through CCTV cameras, crucial for curbing spread of COVID-19. Employed new technologies such as YOLO model and OpenCV methods for object detection to identify individuals in video frames computing 98% accuracy.

Algorithmic paper trading using Deep Reinforcement Learning

Python, FinRL

- Automated real-time creation of trading strategies for stocks such as DOW 30, using Proximal Policy Gradient (PPO) algorithm, and quantitative statistical principles for enhanced forecasting to optimize profits. Deployed resultant strategy via ALPACA finance API, accomplishing an impressive portfolio statistics of \$1,034,450.54 from an initial \$1,000,000 investment during paper trading.

TECHNICAL SKILLS

Programming languages: Python, R, Java, C++, C#, HTML/CSS, JavaScript, SQL

Technologies/Frameworks: Linux, PostgreSQL, MongoDB, MySQL, Scikit Learn, TensorFlow, Selenium, Pandas, NumPy, Keras

Developer Tools: VS Code, AWS, GCP, Azure, Tableau, Power BI, Apache Kafka, Spark, Hive, Docker, Hadoop, Kubernetes

CERTIFICATIONS

- Machine Learning by IBM
- Deep Learning Specialization
- Become a Data analyst specialization
- Spark AR by FACEBOOK