Kathan Joshi

Harrisburg, PA | 240-390-5282 | joshikathan03@gmail.com | linkedin.com/in/kjoshi03/ | AWS AI Certified

Education

Pennsylvania State University (3.63/4.00) — Pennsylvania, USA
Master of Science in Computer Science — *Aug 2023 – May 2025*

Ahmedabad University (3.41/4.00) — Gujarat, India
Bachelor of Technology in Computer Science and Engineering — *Aug 2019 – Jun 2023*

Work Experience

Machine Learning Research Assistant/TA — *Sep 2023 – Present*
Pennsylvania State University — *Pennsylvania, USA*

- Engineered and fine-tuned a suite of 7 ML models (AdaBoost, CatBoost, XGBoost, Random Forest, KAN, MLR, DNN), to predict the LOS of a hospital patient, boosting performance metrics by 12%.
- Implemented a comprehensive five-step data pipeline that enhanced model performance by 38% and reduced inconsistencies by 45%.
- Leveraged Python, SQL, and Dask to analyze HCUP's complex data sets, and Power BI for interactive data visualizations that uncovered trends, boosting predictive accuracy by 25% and predictive insights by 10%.
- Served as a TA for Statistics, Mathematics, Calculus, ML, NLP, & Data Science courses, supporting 60+ students.

Data Scientist — *Aug 2022 – Jul 2023*
Safal Industries — *Gujarat, India*

- Conducted comprehensive Exploratory Data Analysis with Python, Pandas, and AWS SageMaker on equipment and production data, uncovering trends that boosted productivity and trimmed operational costs by 7%.
- Designed and deployed a PyTorch-based computer vision defect detection model that achieved a 92% accuracy rate and reduced manual inspection time by 30%.
- Architected a data science pipeline to embed ML models into production systems, resulting in a 20% faster quality control cycle and a 15% improvement in predictive maintenance.
- Formulated a sentiment analysis solution to extract key issues from client reviews, increasing issue identification efficiency by 25% and reducing churn by 8%.

Data Scientist – Simulation & Visualization Intern — *May 2022 – Aug 2022*
Indian Space Research Organization (ISRO) — *Gujarat, India*

- Built a virtual rover control system in Python using Vizard7 that simulated real-world physics with 20% elevated simulation fidelity.
- Automated predefined path navigation for the rover, reducing manual oversight by 15%.
- Constructed a streamlined data collection framework that integrated rover state data into a local database, enhancing access by 30% and cutting processing delays.

Technical Skills

Programing Languages: Python, SQL, R, Java, JavaScript, HTML/CSS, C/C++
Frameworks: Langchain, Streamlit, TensorFlow Lite, React, Flask
Big Data/ Tools & Software: Git, Docker, AWS, Power BI, Tableau, Gurobi, Hadoop, Spark, Linux
Libraries: Tensorflow, Pytorch, NumPy, Pandas, Dask, SKlearn, NLTK, Spacy, Matplotlib, Gurobipy, seaborn, plotly
Certifications: AWS AI Practitioner, Machine Learning Specialization, GenAI with LLMs (AWS, DeepLearning.AI)

Projects

Sentiment Analyzer | *Pytorch, Numpy, Pandas, SkLearn, transformers, Matplotlib* — *Sep 2024 – Nov 2024*

- Fine-tuned a BERT LLM model using PyTorch on the SMILE Twitter dataset for multi-class sentiment analysis, elevating overall classification performance measurement by 23%.
- Mitigated class imbalance via back translation during preprocessing, aided by data visualization of class distributions, which enhanced accuracy by 2% for under-represented sentiment classes.

Fake News Detector | *Tensorflow, Numpy, Pandas, SkLearn, NLTK* — *Nov 2023 – Dec 2023*

- Developed a lightweight LSTM model (time series model) in TensorFlow for fake news detection, achieving an impressive 99.69% classification accuracy.
- Optimized text preprocessing with NLTK to reduce trainable parameters by 78.57%, further enhancing model accuracy by 2%.