

KSHITIJ JOSHI

+1 (667) 308-1785 ✉ kjosshi15@jhu.edu 🌐 LinkedIn 🐙 GitHub

EDUCATION

The Johns Hopkins University (GPA - 3.72/4)

Aug 2023 – Dec 2024

Master of Science in Data Science

Baltimore, Maryland

Coursework: Data Mining, Machine Translation, Statistical Methods & Data Analysis, Algorithms for Data Science, Intro to Optimization, Software Engineering for Data Science, Equity Markets & Quantitative Trading.

Gujarat State Fertilizers and Chemicals University (GPA - 3.44/4)

Aug 2019 – Jun 2023

Bachelor of Engineering in Computer Science & Engineering with Specialization in Data Science

Vadodara, India

Coursework: Data Structures & Algorithms, Machine Learning for Intelligent Systems, Deep Learning, NLP, Big Data, Computer Networks, Operating Systems, Android App Development, Software Engineering

TECHNICAL SKILLS

Programming Languages and Tools: Python, R, SQL, Julia, Git, GitHub, GitLab, FastAPI, Streamlit

Gen AI, ML & DL Frameworks: TensorFlow, PyTorch, Scikit-Learn, Keras, FastAI, OpenCV

Data Engineering, MLOps, and Cloud Platforms: Apache Spark, Dask, Airflow, Kafka, Docker, Kubernetes, MLflow, TensorFlow Serving, AWS, Google Cloud Platform, Microsoft Azure, Vector Databases

Data Analysis & Visualization: Stata, SPSS, Matplotlib, Seaborn, Plotly, Tableau, Power BI

RELEVANT EXPERIENCE

Johns Hopkins University - Center for Language & Speech Processing

Aug 2024 – Present

Research Assistant

Baltimore, MD

- Compared LLMs and traditional translation for Gujarati to English, improving accuracy by 15% using the Flores200 dataset, highlighting LLMs' strength in low-resource settings.
- Developed and implemented various Machine Translation techniques such as Transformer models and Neural Machine Translation to enhance Gujarati to English translation, focusing on optimizing performance and accuracy.
- Evaluated translation outputs to determine the best approach for low-resource languages, achieving a **20%** reduction in errors with LLMs and providing recommendations to improve translation quality.

Johns Hopkins University - Advanced Robotics (ARCADE) Lab

Jan 2024 – Present

Research Assistant

Baltimore, MD

- Improved **Cognitive load estimation** in Telerobotic surgery by developing deep learning models using Pupillometry data, resulting in significant accuracy improvements across various surgical scenarios.
- Enhanced model robustness for cognitive load detection by applying spectral domain analysis and gaze entropy techniques, reducing error rates by **20%**.
- Streamlined surgical training processes by integrating machine learning algorithms for real-time data processing and visualization, increasing decision-making efficiency and reducing training time by **30%**.

Johns Hopkins University - Bloomberg School of Public Health

Jun 2024 – Aug 2024

Data Science Intern

Baltimore, MD

- Optimized API integration by developing a dynamic endpoint retrieval system using RAG, resulting in a 20% increase in data flow continuity and a 15% improvement in model accuracy and reliability.
- Boosted data collection processes by implementing User-Agent and proxy rotation strategies, maintaining uninterrupted access to diverse web data sources and increasing data availability, crucial for machine learning model training.
- Optimized data scraping efficiency with rate limiting and randomization strategies, extending data collection lifespan, reducing latency by 15%, and ensuring contextually relevant data augmentation for training models.

PROJECTS

CricketMatchPredictor: Machine Learning Models for IPL Outcome Prediction

Spring 2024

- Improved IPL match outcome prediction accuracy by 25% to 58.6% using Data Science techniques such as Markov Chains, Bi-LSTM networks, and Ensemble methods (Gradient Boosting, XGBoost), enhancing sports analytics capabilities.

LegalEase: AI-Driven Legal Document Translation

Fall 2023

- Achieved a BLEU score **0.58** for complex legal text translation by **fine-tuning the IndicTrans model with Fairseq**, enhancing the model's capability to handle linguistic complexities and reducing manual translation time by 40%.

Finland Healthcare Analysis

Fall 2023

- Analyzed Finland's healthcare data with Difference in Difference, Linear Regression, and **Causal Inference**, revealing a 30% correlation with improved services and a **25%** quality boost, using visualizations for policy recommendations.