ISHANI ARYA

iarya1@jh.edu | +1 (443) 226-0240 | linkedin.com/in/ishani-arya | github.com/ishani2202 | https://ishani-arya.netlify.app/

EDUCATION

JOHNS HOPKINS UNIVERSITY | Baltimore, MD
Master of Science in Engineering in Data Science | Aug 2024 - May 2026
Relevant Courses: Data Science, Machine Translation, Statistical Modelling, ML for Healthcare, NLP: Self Supervised Models

LM THAPAR SCHOOL OF MANAGEMENT | Patiala, India | CGPA: 3.42/4
Masters of Business Administration | Aug 2023 - June 2024
Relevant courses: Financial Derivatives & Risk Management, Options Pricing, Portfolio Management, Financial Regulations & Norms

THAPAR UNIVERSITY | Patiala, India | CGPA: 3.20/4
Bachelors of Technology in Computer Science| Aug 2019 - June 2023
Relevant courses: Deep Learning, Mathematics in ML, Probability and Statistics, Data Structures, Design and Analysis of Algorithm

EXPERIENCE

Research Assistant: Financial Data Science | **Johns Hopkins University** | Baltimore, USA | Nov 2024 - Present
- Engineered an **ETL pipeline** in **Python** and **Stata** for analyzing unstructured SEC filings (10-K & 10-Q) from BDCs
- Applied data **wrangling**, **feature engineering**, and **regex** focused classification to convert complex financial data into a machine-readable format, preparing the dataset for **predictive modeling** and financial time-series forecasting using **ARIMA**

Data Science Trainee | **LG Electronics** | Noida, India | Feb 2024 – July 2024
- Designed & implemented **web scraping pipelines** using **Selenium and Beautiful Soup** to extract **Voice of Customer** data from YouTube and Flipkart, enabling actionable insights that enhanced product development strategies
- **Fine-tuned BERT** for sentiment analysis and utilized **Latent Ditrichet Allocation** (LDA) for topic modeling on customer feedback

Big Data Intern | Reliance Jio | Mumbai, India | June 2023 – Aug 2023
- Developed a scalable **ETL pipeline using Apache Spark** to process **1M+ tweets**, applying **Naive Bayes** for sentiment classification and **SVM** for engagement prediction, and visualized insights in Tableau, boosting campaign effectiveness by **15%**

Machine Learning Researcher| **Thapar University** | Patiala, India | June 2022 – May 2023
- Developed a deep learning-based **waste classification boat** using **ResNet and TensorFlow**, achieving **93.07% accuracy** in real-time classification of biodegradable vs. non-biodegradable waste, deployed on **Raspberry Pi** for efficient edge processing
- Engineered **real-time image classification** using TensorFlow for live camera input, integrating **GPIO and Pygame** for image capture and processing, resulting in a **published patent** for the autonomous system

TECHNICAL SKILLS & CERTIFICATIONS

- **Programming Languages:** Python, R, SQL, C++, CSS, HTML, Java
- **Tools, technologies, frameworks:** Scikit-learn, Pandas, NumPy, Matplotlib, NLTK, TensorFlow, PyTorch, Keras, OpenCV, Spark, Hadoop, Kafka, CUDA, Flutter, ReactJS, Streamlit, Flask, AWS, Git, Linux, MS Excel, Docker, Azure, GCP, Jenkins, Lang chain, LLMs
- **Databases & BI tools:** MySQL, PostgreSQL, MongoDB, AWS(S3), GCP, Tableau, Power BI, Grafana.
- **Certifications:** Data Science and Deep Learning Specialization (Coursera), LangChain Chat with Your Data Project (Coursera)

PROJECTS

MelanoViT | [Github]
- Developed a robust **melanoma classification** framework using **Vision Transformers (ViT) and DinoV2**, achieving state-of-the-art **accuracy of 99.03%** with metadata integration, **surpassing all previous CNN-based models**.
- Enhanced model sensitivity by addressing class imbalance with weighted loss functions and improving generalization using **augmentation techniques** like hair removal and geometric transformations

Pandemic Tracker | [Github] | [Interactive Website]
- Developed a COVID-19 global dashboard using **Python, Pandas, Plotly**, and **time-series analysis** techniques such as **rolling averages** and trend analysis to visualize and monitor pandemic data.
- **Deployed** the application on **Streamlit Cloud** with dynamic visualizations (choropleths, bar graphs, line plots) and interactive filters, enabling in-depth exploration of COVID-19 trends and insights.

EchoTranslate | [Github]
- Designed a **Transformer-based pipeline** for **ASR** and **multilingual translation**, using doctor-patient conversations aimed at **automating medical transcription**
- Leveraged **Torchaudio** and **Librosa** for spectrogram feature extraction, signal denoising, & preprocessing. Fine-tuned on a domain-specific medical corpus achieving a **5.8% BLEU score improvement** and **12.5% WER reduction**