Aziz Ashfak

01730644634 | azizashfak@gmail.com | github | portfolio | linkedIn | kaggle

PROFILE

AI Engineer and **researcher** skilled in developing and deploying **machine learning**, **deep learning**, and **generative AI** solutions.

EDUCATION

Noakhali Science and Technology University *Noakhali,Bangladesh*

Bachelor of Science in Statistics *Apr* 2022 – Aug 2026

EXPERIENCE

Data Analyst (Research Collaboration)

Hestia Traffic App Project —Collaboration with PhD Researcher, Northern Arizona University Jan 2025 – Feb 2026

- Analyzed large-scale traffic and road accident datasets to identify key factors influencing crash severity and transportation safety outcomes.
- Applied core **data analysis techniques**, including data cleaning, preprocessing, and exploratory data analysis, to uncover patterns in traffic violations, accident frequency, and environmental conditions.
- Conducted statistical analysis to investigate influential variables affecting road safety and accident risk.
- Contributed to the development and analytical evaluation of the **Hestia Traffic App**, a data-driven platform designed to monitor traffic conditions and analyze accident patterns.
- Utilized Python-based data analysis tools (Pandas, NumPy, Matplotlib) to process, analyze, and visualize complex transportation datasets.
- Collaborated with a research team to translate analytical insights from traffic data into actionable findings supporting transportation safety research.

AI Engineer (remote, full-time**) | ROYALX LLC ,** Dhaka, Bangladesh *Feb 2025 – Sep 2025*

• Developed and deployed **large language model (LLM)** applications using **OpenAI**, **Groq**, **Hugging Face models**, **LangChain**, and **LlamaIndex** for text generation, summarization, chatbots etc.

- Built **retrieval-augmented generation (RAG) pipelines**, including **multimoda**l inputs, improving accuracy and efficiency of AI-powered solutions.
- Fine-tuned transformer models for tasks including question-answering, classification, and document understanding.
- Designed and deployed AI agents capable of reasoning, multi-step problem-solving, and autonomous task execution using LangChain workflows.
- **Implemented production-ready AI systems** using **Flask**, **FastAPI**, **Streamlit**, and **cloud services**, ensuring scalability, robustness, and low latency.
- Collaborated with remote teams to deliver client-focused AI solutions, optimizing performance and usability. **Key Achievement**: Successfully integrated **LLMs**, **LangChain workflows**, **LlamaIndex**, and **AI agents** into enterprise applications, resulting in faster automation, improved decision-making, and enhanced user experience.

AI Tutor (remote, part-time**)** | [**Ai Devlop** Dhaka, Bangladesh *Apr 2025 – Aug 2025*

- Taught **100+ students** the foundations of **Exploratory Data Analysis (EDA), Statistical Analysis, Machine Learning, and Deep Learning**.
- Designed and delivered structured lessons, practical coding sessions, and project-based learning.
- Guided students on implementing **end-to-end ML pipelines**, from data preprocessing to model deployment.
- Mentored students in developing AI projects, improving their problem-solving and research skills.

Key Achievement: Successfully trained and mentored a diverse group of learners, many of whom progressed to building their own **ML/DL research projects and applications**.

RESEARCH WORK

Optimizing U-Net for Human Segmentation – Evaluating state-of-the-art encoders (ResNet, EfficientNet) and optimizers (AdamW, SGD) to enhance segmentation accuracy on a human segmentation dataset, proposed for ICRA 2026. (**Proposed**)

Evaluating Multimodal StatTransformer Architectures with Large Language Models for Statistical Reasoning and Real-World Data Integration – Evaluating Multimodal StatTransformer Architectures with Large Language Models for Statistical Reasoning and Real-World Data Integration (**In progress**)

PROJECTS

Brain Tumor Detection [project-link] | [web-sevice-link] | [demo run]

- Built a 98.52% accurate Brain Tumor Detection system using transfer learning with Lenet, ResNet152V2 and DenseNet121 on a Kaggle MRI dataset, optimized with normalization and augmentation.
- Deployed a scalable ONNX-based web app on Render for fast brain tumor predictions.
- Demonstrated expertise in transfer learning, deep learning, and CI/CD deployment.

Indian Flight Price Prediction [project-link] | [web-service-link] | [demo run]

- Engineered a 99.69% accurate flight price prediction model using BaggingRegressor on a 300,261-entry Easemytrip dataset, leveraging MinMaxScaler, OneHotEncoder, and SimpleImputer.
- Deployed a scalable web app on Render with CI/CD via GitHub Actions for real-time flight price predictions.
- Demonstrated expertise in data preprocessing, model optimization, and efficient ML pipeline deployment.

Thyroid Disease Prediction [project-link] | [web-service-link] | [demo run]

- Developed a 98.7% accurate Thyroid Disease Prediction system using RandomForestClassifier, leveraging MinMaxScaler, OneHotEncoder, and SimpleImputer on UCI dataset.
- Engineered a scalable ML pipeline with hyperparameter tuning and deployed a on Render for real-time predictions.
- Demonstrated expertise in data preprocessing, model optimization, and CI/CD integration with GitHub Actions.

Multimodal Research Paper Analyst [project-link] | [demo run]

- Developed a state-of-the-art Multimodal RAG system for research paper analysis, leveraging Groq and Llama to summarize content and interpret graphs/images with high precision.
- Deployed a Streamlit web app with modular pipelines, enabling custom queries and structured report exports in Word, PDF, CSV, and Markdown formats.
- Showcased expertise in multimodal LLMs, prompt engineering, and scalable deployment for impactful academic and research applications.

QnA RAG Bot [project-link] | [web-service-link] | [demo run]

- Engineered a cutting-edge RAG-powered Q&A bot using FastAPI, LangChain, and HuggingFace, delivering precise context-aware answers from PDFs with advanced figure/graph analysis.
- Deployed a scalable, modular web app on Render, supporting multi-format exports (Markdown, CSV, DOCX, PDF) and FAISS-driven vector search.
- Demonstrated expertise in LLM integration and robust pipeline design for impactful research applications.

Face Mask Detection [project-link] | [demo run]

- Pioneered a high-accuracy Face Mask Detection system using YOLO and transfer learning, enhancing public health by enabling real-time monitoring of mask compliance on a Kaggle dataset of 853 images.
- Deployed an accessible Flask web app with ONNX for optimized inference, scalable for real-world applications like workplaces and public transport.

Technical Skills

Programming Languages: Python, R, SQL, C(basic), Java(basic)
DSA : Arrays,Strings ,Hash Maps, Linked Lists, Trees, Graphs, Recursion, Sorting/Search, Dynamic Programming
Frameworks & Libraries: PyTorch, TensorFlow, Keras, Scikit-learn, OpenCV, LangChain, Hugging Face
GenAI & LLMs: OpenAI API, Hugging Face Transformers, Groq API, LangChain
Web Development & Deployment: FastAPI, Streamlit, Flask, Gradio
Vector Databases & Retrieval: ChromaDB, FAISS, Pinecone
CI/CD & DevOps: GitHub Actions, Git
Cloud Platforms & Hosting: Google Colab, Render, Kaggle
Automation: OCR, Multimodal Pipelines, RAG Pipelines, Agentic-RAG, n8n, crewAI, Langgraph **Data Processing & Analysis:** Pandas, NumPy

ACHIEVEMENTS

- Optimizing U-Net Performance: Evaluating State-of-the-Art Encoders and Optimizers on a Human Segmentation Dataset Proposed for ICRA 2026
- Developed and deployed modular GenAI-powered systems for research automation.