Koshik Debanath

+8801855675763 | koshik.debanath@gmail.com | LinkedIn | GitHub | Google Scholar | Portfolio

Education

Rajshahi University of Engineering & Technology — Rajshahi, Bangladesh
B.Sc. in Computer Science and Engineering; ***CGPA: 3.27/4.00*** — *Jan 2018 – Sep 2023*

Skills

Programming: C/C++, Python, Java, MATLAB, JavaScript

Tools & Platforms: Git, Docker, CI/CD, MLOps, MySQL, MongoDB, SQLite, Django, Flask, FastAPI

AI/ML: PyTorch, TensorFlow, scikit-learn, OpenCV, NLP (LLMs, Fine Tuning, Prompt Engineering), Computer Vision, Time Series Analysis, Generative AI, Data Mining

Publications

- **K. D. Nath**, A. F. M. M. Rahman and M. A. Hossain, *"An Attention-Based Deep Learning Approach to Knee Injury Classification from MRI Images,"* 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ICCIT60459.2023.10441340.
- **K. Debanath**, S. Aich and A. Y. Srizon, *"Advancing Low-Resource NLP: Contextual Question Answering for Bengali Language Using Llama,"* 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2025, pp. 1-6, doi: 10.1109/ECCE64574.2025.11013841.
- S. Aich, **K. Debanath** and A. Y. Srizon, *"Distinguishing Between Formal and Colloquial: A Multilingual BERT Approach to Bengali Language Classification,"* 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2025, pp. 1-6, doi: 10.1109/ECCE64574.2025.11013999
- **K. Debanath**, S. Aich and A. Y. Srizon, *"Analyzing Bot Activity and Political Discourse in the 2024 U.S. Presidential Election: A Machine Learning Approach to Misinformation and Manipulation,"* **Accepted**, To appear in 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM-2025).
- S. Aich, **K. Debanath**, and A. Y. Srizon, *"Distinguishing Human-Written and AI-Generated Text: A Comprehensive Study Using Explainable Artificial Intelligence in Text Classification,"* **Accepted**, To appear in 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM-2025).

Experience

Manaknightdigital Inc. — Toronto, Ontario, Canada (Remote)
Data Scientist — *Mar 2023 – Apr 2025*

- **Chatbot Development**:
  - Collected and processed product information using Excel, pandas, and openpyxl.
  - Integrated GPT-4 to respond to user queries and manage token size limitations.
  - Utilized libraries like nltk, sklearn, and Flask for deploying the chatbot.
- **Fraud Detection System**:
  - Performed EDA and feature extraction on transaction datasets.
  - Developed and optimized ML models including Xgboost, SVC, and Logistic Regression.
  - Achieved 90% accuracy in detecting fraudulent transactions and deployed the system using Flask.
- **Data-driven ChatBot for Financial Queries**:
  - Implemented RAG and Pinecone, enhancing data retrieval speed by 40%, enabling faster decision-making for lenders.
  - Improved data retrieval accuracy by 25%
  - Applied Beautiful Soup and PyPDF2 for data scraping and processing.
- **Sports Data Analysis ChatBot**:
  - Scraped and analyzed football data to predict match outcomes.

    - Integrated RAG and Pinecone for efficient data querying and vector database management.
    - Employed Beautiful Soup and PyPDF2 for data collection, analyzing 2 million football data points to achieve a 90% prediction accuracy, supporting strategic betting decisions.
- **Custom Image Generation System**:
    - Developed an image generation platform using Stable Diffusion.
    - Fine-tuned custom models to generate images based on user-defined presets.
    - Utilized PyTorch and transformers for model training and deployment and finally used Docker for containerization.
- **AI-driven Data Matching System**:
    - Organizational data was segmented using models such as Llama-2-7B and then fine-tuned to extract sections and subsections.
    - Applied cosine similarity for matching data to specific tenders.
    - Integrated GPT-4 for generating rationale from corresponding data.
    - Matched organizational data against specific tenders, increasing successful tender submissions by 70%.
- **AI-Powered Collectible Authentication & Appraisal Platform**:
    - Trained deep learning models (PyTorch/TensorFlow, e.g., InceptionV3, ResNet50, CLIP) for image classification (authenticity) and similarity search.
    - Engineered an efficient CLIP+FAISS image similarity system for large-scale appraisal lookups.
    - Developed Flask/FastAPI APIs to serve model predictions (classification, similarity, appraisal).
    - Designed a multi-modal tag identification system using Serverless (RunPod API), TF-IDF, and CLIP/FAISS similarity.
    - Implemented asynchronous data pipelines (aiohttp, asyncio, pandas) for large-scale image and metadata ingestion from APIs.
    - Developed a Streamlit web application for user image uploads and displaying similarity/appraisal results via API calls.

Universal Machine Inc. Sunnyvale, California, United States (Remote/Hybrid)
Software Engineer - I *Apr 2025 – **Present***

- **YouTube Live Stream Bot**:
    - Developed Chrome Extension automating YouTube Live chat using JavaScript, Chrome APIs, and async requests.
    - Integrated YouTube & OpenAI APIs for real-time chat fetching/posting and AI response generation.
    - Engineered AI features managing conversational history (chrome.storage) and prompt engineering for context/recall.
    - Implemented secure Google OAuth (chrome.identity) and robust error handling for external APIs.

Competitions

Hackathon Champion at Machine Hack: Global Ranking: 539 Out Of 8861

Rental Bikes Volume Prediction: Rank: 3rd

News Category Prediction: Rank: 7th

Data Science Student Championship: Secured 7th position among 1029 participants from engineering colleges and universities across India in jointly hosted by MachineHack Generative AI and Praxis Tech School

LLM Hackathon: Decoding Discourse - AI vs Human: Rank: 5th Out of 227.

Predicting House Prices in Bengaluru: 24th Rank Out Of 2885 with Accuracy of 87%.

Subscriber Prediction Talent Search Hackathon: Rank: 26th Out Of 5045.

Analytics Olympiad 2022: Rank: 82 Out Of 1029.

Data Science Student Championship - South Zone: Rank: 73rd Out of 554.

Decoding Discourse - AI vs Human: Rank: 5th Out of 293.

Projects

Bangladesh Stock Price Forecast | *App*

- Goal: Forecasting Stock Price of Bangladesh by using LSTM (Long Short-Term Memory) networks
- Library: pandas, numpy, matplotlib, bdshare, keras, streamlit, and plotly

UberRidePrediction | *PyPi* | *WebApp*

- Goal: UberRidePrediction is a Python module designed to predict Uber ride prices based on factors like location coordinates, number of passengers, and ride time using machine learning algorithms(Xgboost).
- Library: scikit-learn, CI/CD Pipeline, FastAPI

PineconeUtils | *PyPi*

- Goal: PineconeUtils is a Python module designed to handle and process data for embedding and indexing using Pinecone, Cohere, and OpenAI services for applications involving text embedding and retrieval augmented systems(RAG) Library: PineconePDFExtractor, openai,cohere,pinecone

Decoding AI vs Human | *WebApp*

- Goal: Decoding AI vs Human is an interactive web application that allows users to put any text and see if a human or an AI wrote it. This application is trained on the Machine Hack dataset. Library/Technology: scikit-learn, AWS

PineconePDFExtractor | *PyPi*

- Goal: PineconePDFExtractor is a Python library for extracting text from PDF files for pinecone. Library: PyPdf2

DataSciencePilot | *GitHub*

- Goal: It is a chat-based interface designed to interact with custom PDF files. It leverages the power of Pinecone for efficient vector database management and LLaMA-2 for advanced query response capabilities Library: Pinecone, Langchain, Transformers

CVAnalyzerPro | *StreamlitApp*

- Goal: matches participant's CVs with the company's requirements and gives scoring Library: openai, Gemini, Streamlit

CaptionCraft | *StreamlitApp*

- Goal: generate caption using Google Gemini API Library: Gemini, Streamlit

PredictStock | *StreamlitApp*

- Goal: to predict the stock of any company like Google, Microsoft, Apple Solution: used LSTM to train the model Library: Tensorflow, Streamlit

Diabetes Prediction | *GitHub*

- Goal: predict whether any patient has diabetes or not Solution: used Artificial Neural Network(ANN) to train the model and predict the disease Library: PyTorch, Flask, Gunicorn

Movie Recommendation | *WebApp*

- Goal: recommend the movie based on the movie entered by user Solution: Used KNN to find the nearest 5 movies using cosine similarity Library: pandas, numpy,sklearn, Flask, scipy

Market Price Prediction | *GitHub*

- Goal: to predict the price of the product using ARIMA, SARIMAX,LSTM, FbProphet, GRU, Xgboost

Potato Disease Classification Using CNN | *GitHub*

- Goal: To classify disease in Potato Solution: the dataset is taken from Kaggle which contains 1506 images with 3 classes Result: Overall accuracy is 100% Library: Tensorflow

Open Source Contribution

- Contributed on openllmetry: Open-source observability for your LLM application, based on OpenTelemetry Contribution

- Contributed on Pinecone Canopy: Retrieval Augmented Generation (RAG) framework and context engine powered by Pinecone Contribution