

Kowshik Deb Nath

☎ +8801855675763 | @ kowshikcse17@gmail.com | 🔗 LinkedIn | 🐙 GitHub | 📁 Portfolio | 📍 Rajshahi, Bangladesh

EDUCATION

Rajshahi University of Engineering & Technology
B.Sc. in Computer Science and Engineering; CGPA: 3.27/4.00

Rajshahi, Bangladesh
Jan 2018 – Sep 2023

SKILLS

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

Databases: MySQL, MongoDB, SQLite

Web Frameworks: Django, Flask, FastAPI, JavaScript

Technologies: Git, Docker, CI/CD Pipeline

Machine Learning/AI: scikit-learn, OpenCV, PyTorch, TensorFlow, MLOps, NLP, Computer Vision, Time Series Analysis, Fine Tuning LLM, Generative AI, Data Mining, Prompt Engineering

EXPERIENCE

Manaknightdigital

Data Scientist

Toronto, Ontario, Canada (Remote)

Mar 2023 – Present, Full-time

- **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 BeautifulSoup 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 BeautifulSoup 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%.

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, Unicorn

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](#)

PUBLICATIONS

- "An Attention-Based Deep Learning Approach to Knee Injury Classification from MRI Images" [Accepted by IEEE Xplore](#)