Kowshik Deb Nath

☐ +8801855675763 | @ kowshik.debanath@gmail.com | in LinkedIn | ⌾ GitHub | ⊕ Portfolio | ⚲ Rajshahi, Bangladesh

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

Languages: C/C++, Python, Java, JavaScript, SQL, MATLAB

Technologies: Flask, MySQL, Git, Docker, AWS, CI/CD Pipeline, OpenCV, PyTorch, TensorFlow, Google Dialogflow, MLOps, NLP, Computer Vision, Time Series Analysis, Fine Tuning LLM's, Generative Adversarial Networks (GAN), Data Mining, Prompt Engineering

EXPERIENCE

Manaknightdigital Toronto, Ontario, Canada (Remote

Data Scientist *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 to retrieve and manage lender data.
    * Enhanced data retrieval accuracy using Cohere reranking.
    * 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 effective data querying.
    * Employed Beautiful Soup and PyPDF2 for data collection.
- **Custom Image Generation System**:
    * Developed an image generation platform using Stable Diffusion.
    * Trained custom models to generate images based on user-defined presets.
    * Utilized PyTorch and transformers for model training and deployment.
- **AI-driven Data Matching System**:
    * Chunked organizational data into sections using models like Llama-2-7B.
    * Applied cosine similarity for matching data to specific tenders.
    * Integrated GPT-4 for generating insights from matched data.

COMPETITIONS

MASTER at Machine Hack: Global Ranking: 310 Out Of 8052, Total Points: 1321.

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

Decoding AI vs Human | *AWS*
- 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 | *GitHub*
- 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 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