Kowshik Deb Nath

✉ kowshikcse17@gmail.com   🌐 kowshik24.github.io/kowshik.github.io/   in linkedin.com/in/kowshik24

⌗ github.com/kowshik24   kaggle.com/kowshikdebnath

EDUCATION

Rajshahi University of Engineering & Technology Rajshahi, Bangladesh
B.Sc. in Computer Science & Engineering;
CGPA: 3.26/4.00 Jan 2018 − Sep 2023

WORK EXPERIENCE

Data Scientist — Manaknightdigital (Toronto, Ontario, Canada) March 22, 2023 – Present, Full-Time

Atlas Copco Chatbot:
Primary Goal: to create a chatbot that will answer the questions based on the requirements using the given data. **Solution:** Collect the data(excel file) which contains various product info(Model Name, Price, Capacity, Products Links). Then used the GPT(GPT- 4) model to answer the questions of what the user need. Also, tackle the token size in the GPT models. **Library:** openai, Flask, openpyxl, pandas **Website**: Atlas Copco

Fraud Transaction Detection:
Primary Goal: Given a dataset of past transactions then from this data set have to build an ML model that can detect fraud transactions. **Solution:** perform EDA and Feature Extraction apply multiple ML algo(Xgboost,SVC,Logistic Regression) then apply hyperparameter optimization and deploy on the server for production. Got 90% accuracy using Xgboost. **Library:** scikit-learn,Flask

Michel AI Parabroker ChatBot:
Primary Goal: scrape data from different lenders then get answer out the data from user query **Solution:** used Retrieval Augmented Generation(RAG) based searching techniques to get the answers and for data storing used Pinecone also used Cohere Reranking to refine the RAG search **Library:** PyPDF2,beautifulsoup,gunicorn,GPT-4, Pinecone,RAG, Cohere, Flask **Website**: Cynario

Football Analyst ChatBot:
Primary Goal: to make a Football Analyst which will analyze the pass data for a particular Team and Player and Predict the future match **Solution:** scrap data from different football data sources and then used RAG,Pinecone and Cohere to get the response out of the huge amount of data **Library:** BeautifulSoup, PyPDF2, RAG, Pinecone, Cohere **Website**: Kaizenwin

Image Generation (Aieventbooth):
Primary Goal: user will add some presets for a given image then it will generate the image **Solution:** used Stable Diffusion Image Generation, trained on custom image and then based on the used presets generates the image **Library:** transformers,PyTorch **Website**: Aieventbooth

Ai Energy:
Primary Goal: clients have some organization's privacy and policy, now a tender will came and we have look up the organization's data how many matches each of the tender's and why matched **Solution:** to solve this problem first chuck the data into different sections for this finetuned different models like Llama-2-7B and 13B also Mistral-7B then used the cosine similarity to get the matched data and then used the GPT-4 to get the answer out of the data **Library:** transformers,PyTorch **Website:** Ai Energy

COMPETITIONS

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

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

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

4. Subscriber Prediction Talent Search Hackathon: Rank: 26th Out Of 504

5. Analytics Olympiad 2022: Rank: 82 Out Of 1029

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

SKILLS

Machine Learning, Deep Learning, Data Analysis & Visualization, GAN, Time Series Analysis, Computer Vision, Cloud Computing, RNN, Prompt Engineering, Fine Tuning LLM's, Statistical Tools, Scripting Language: Python, HTML, C++, C, Java, JavaScript, Database: MySQL, SQLite, NLP, Text Normalization, BERT, Sentiment Analysis, Word Embedding, Data Visualization & Analysis: MS Excel, PowerBI, matplotlib, Web Framework: Flask, Libraries: Tensorflow, Scikit-learn, Seaborn, PyTorch

PROJECTS

1. **Potato Disease Classification Using CNN: Goal:** To classify disease in Potato **Solution:** dataset is taken from Kaggle which contains 1506 images with 3 classes **Result**: Overall accuracy is 100% **Library:** Tensorflow ◯Potato-Disease-Classification

2. **Market Price Prediction: Goal:** to predict the price of the product using ARIMA, SARIMAX, LSTM, FbProphet, GRU, Xgboost ◯ Market Price Prediction

3. **Movie Recommendation: Goal:** recommend 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 ◯ Movie Recommendation

4. **Diabetes Prediction: Goal:** predict whether any patient have diabetes or not **Solution:** used Artificial Neural Network(ANN) to train the model and predict the disease **Library:** PyTorch, Flask , Gunicorn ◯ Diabetes Prediction

5. **PredictStock: Goal:** to predict the stock of any company like Google, Microsoft , Apple **Solution:** used LSTM to train the model **Library:** Tensorflow,Streamlit **StreamlitApp:** PredictStock

6. **CaptionCraft: Goal:** generate caption using Google gemini API **Library:** Gemini, Streamlit **StramlitApp:** CaptionCraft

7. **CVAnalyzerPro: Goal:** matches participants CV with company's requirements and give scoring **Library:** openai, gemini , Streamlit **StreamlitApp:** CVAnalyzerPro

8. **DataSciencePilot: 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 ◯ DataSciencePilot

PUBLICATIONS

1. Title: "An Attention-Based Deep Learning Approach to Knee Injury Classification from MRI Images" Accepted by IEEE Xplorer at ICCIT-2023.