Harsh Vardhan Singh Chauhan

+91-9953944291 | medium.com/@harshtherocking | linkedin.com/in/harshvschauhan | https://github.com/Harshtherocking

EXPERIENCE

Data Miner   Dec. 2023 – Jan 2024
ScimerTech   *Remote*
- Web scrapping of data from **NIST web-book** improving the traditional data collection process by **85%**
- **Spectroscopic data analysis** of transmittance and absorbance of food sample for extraction and identification of chemical compounds and functional groups
- **Noise reduction** and smoothing algorithms used to pre-process spectral data to enhance the spectra quality, improving accuracy upto **79%**. **1-D Convolution** & **Random Forest** algorithms used for feature extraction and improving the improving the precision of spectroscopic analysis

Data Science Intern   July. 2023 – Aug, 2023
Oasis Infobyte   *Remote*
- Prediction model optimization for Car Price Regression models and Spam Email Classification along with Unemployment Analysis using **Principal component analysis** & **Feature extraction**.
- **GUI implementation** for enhancing user consuming experience and seamless interaction with the software

PROJECTS

PyTorch Chatbot | *open source contribution*   Sept 2024 – Present
- Constructed a **Vector Database** on **Milvus** to store and retrieve semantic embeddings of textual chunks.
- Implementing **Retrieval Augmented Generation (RAG)** using **Llama Index** to enhance model performance by **dynamically retrieving and incorporating** relevant information from website knowledge bases.
- Employing advanced **semantic retrieval techniques** to extract relevant information from vector database.
- Optimizing **language model prompts** and weights and **monitoring** with **DSPy**

Graphical Sentimental Analysis   May 2024 – Present
- Developed **Graph** based model to train **Edge Embedding** representing syntactic dependencies in sentences.
- Implemented **Message Passing** using **Graph Convolution Network** in **PyTorch Geometric** to utilise words, **Node** embedding and dependencies, **Edge** embedding.
- Leveraging the **parallel computation** capabilities of **Transformer**-based networks for efficient training.

IIRS Spectral Classification of Lunar Surface | *Bhartiya Antariksh Hackathon, ISRO*   July 2024 – Aug 2024
- Utilized MATLAB to visualize and analyze **IIRS hyperspectral image** enabling lunar geological analysis.
- **Wavelet Transformation** for noise reduction to enhance the quality and interpretability of hyperspectral data.
- Developed spectral classification models using **KNN** and **Gaussian clustering** attaining **82'% accuracy** with successful GUI implementation via **MATLAB App Designer**.

Vue - Video Streaming Platform   Oct. 2023 – Nov. 2023
- Anime streaming platform with dynamic front-end, **login system**, **hosting room** and **comment** features.
- Employed **Torch Vision** for pre-processing of profile pictures and utilized **OpenCV** for **Face Detection** and **Background Removal** algorithms to isolate and refine facial features.
- Utilising profile image to deliver custom **Cartoon Face** using **Hugging Face** pre-trained model.

TECHNICAL SKILLS

Languages: Python, C/C++, SQL, Java, Javascript, HTML, CSS, PHP, Latex, Bash, Markdown
Frameworks: FastAPI, Pytorch, TensorFLow, PyGeometric, Hugging-Face, LangChain, LlamaIndex, StreamLit
Developer Tools: Git, Docker, MongoDB, MLFlow, MATLAB, Envi, ImageJ
Libraries: Pandas, NumPy, Matplotlib, Keras, SciPy, NLTK, Seaborn, OpenCV, Scikit-Learn, Spacy, NetworkX, beautifulSoup, Scrapy, OpenNLP, Graph4NLP, Fasttext, Transformer, Accelerate, PEFT, Milvus, DSPy

ACHIEVEMENT

Bhartiya Antariksh Hackathon ISRO   NRSC Hyderabad, India
1st Runner-up : *IIRS Spectral Image Classification for the Lunar Surface using AI/ML*   *Aug 2024*

EDUCATION

Indian Institute of Information Technology, Kottayam   Kerala, India
BTech Computer Science, (Artificial Intelligence and Data Science) : 8.3 CGPA   *Nov. 2022 – Present*