KSHITIZ AGRAHARI

DELHI, INDIA

EDUCATION

Madan Moahan Malaviya University of Technology, Gorakhpur

B. Tech - Computer Science and Engineering - CGPA - 8.32

EXPERIENCE

Tata Consultancy Services 🗹

Machine Learning Engineer

- Spearheaded the development of a machine learning model that stratified customers into 5 distinct segments, leveraging over 100,000+ purchasing records, which catalyzed a 30% surge in customer engagement.
- Executed comprehensive data cleansing and normalization processes, addressing 5% missing values and 3% outliers, and engineered 10+ key features, resulting in a 15% enhancement in model accuracy.
- Conducted in-depth Exploratory Data Analysis (EDA) on a dataset comprising 1 million data points. uncovering 3 major trends and 5 significant correlations using advanced **visualization** techniques such as histograms, box plots, and scatter plots.
- Applied a suite of clustering algorithms, including K-Means, Hierarchical Clustering, and DB-SCAN, on a dataset of 100,000+ customer records, achieving a 25% enhancement in clustering precision.
- Amplified conversion rates by 25% and elevated the average order value by 20% through meticulously targeted marketing strategies derived from model-driven insights.
- Reduced customer churn by 15% and enhanced Customer Lifetime Value (CLV) by 20% through the deployment of data-driven retention methodologies.

TECHNICAL SKILLS

LANGUAGES: Python, C/C++, SQL

CORE: Machine Learning, Natural Language Processing(**NLP**), Artificial Intelligence, Data Visualization, Deep Learning, Data Science, Computer Vision, Deep Neural Networks, Transformers, Large Language Models(LLMs), LSTM, Bi-LSTM, Encoder-Decoder, FastText, BERT

FRAMEWORKS/LIBRARIES: Numpy, Pandas, Scikit Learn, PyTorch, TensorFlow, LangChain, HuggingFace, YOLO, openCV

TOOLS: VS Code, **Docker**, Render, Git & Github, Jupyter Notebook, Google Colab, Kaggle, Streamlit, Roboflow, LabelImg

PROJECTS

Virtual Interview System 🗗

November 2024 - January 2025

- Technology: LLM, LangChain, Whisper, FAISS, Streamlit
- Designed and developed an AI-powered Virtual Interview Application using Streamlit, LangChain, and GrogLLM to conduct automated technical interviews.
- Integrated GPT-based Large Language Models (LLMs) to dynamically generate context-aware, domain-specific interview questions using **PyPDF2** for candidate resume parsing and **FAISS** for vector embeddings.
- Implemented speech synthesis using **gTTS** (Google Text-to-Speech) to convert interview questions into audio, enhancing accessibility.
- Utilized Whisper (OpenAI) for high-accuracy Speech-to-Text (STT) conversion, enabling precise voice-to-text transcription of candidate responses
- Implemented **timer-based** interview workflow, session state management in Streamlit to track interview progress, question playback, and response submission.
- Integrated LangChain Memory to maintain context in multi-turn interviews, allowing follow-up questions based on previous responses

2019 - 2023

December 2023 - Present Delhi. India

Gorakhpur, Uttar Pradesh

• Incorporated **JSON-based response** storage for **structured data collection** and efficient processing of candidate answers.

Real Time Face-Mask Detection Web App

September 2024 - November 2024

March 2024 - June 2024

- Technology: YOLOv8, OpenCV, Roboflow, Ultralytics, Transfer Learning
- Designed and implemented a **real-time face mask detection** system using YOLOv8 and OpenCV, capable of detecting masked and unmasked faces in **live video streams** with high accuracy.
- Created a high-quality labeled dataset by **manually annotating** images using **Roboflow**, ensuring precise bounding box placements and well-defined class distributions for effective model training.
- Trained a custom YOLOv8 for **100** epochs, achieving Precision: **83.4%**, Recall: **79.3%**, and mAP: **84.4%**, optimizing model performance for real-world scenarios.
- Integrated OpenCV for real-time frame processing, face detection, and overlaying detection results with bounding boxes and labels.
- Implemented **data augmentation** techniques, such as flipping, rotation, and brightness adjustments, to improve model generalization across various lighting conditions and perspectives.
- Fine-tuned the YOLOv8 model using transfer learning, leveraging **pre-trained weights** to improve detection performance on the custom face mask dataset.
- Developed a **scalable framework** that can be further extended for deployment on edge devices, **cloud platforms**, or integrated into real-world surveillance systems for public safety applications.

Multi-Lingual Translator Web App

- Technology: Python, PyTorch, Hugging Face Transformers, SacreBLEU, Torch DataLoader, AdamW Optimizer, Streamlit
- Built a multilingual translation app using **Streamlit** and **Hugging Face's NLLB-200 model**.
- Translated English text into Hindi, Gujarati, and Marathi with high accuracy.
- Implemented $\mathbf{BLEU},\,\mathbf{chrF},\,\mathrm{and}\,\,\mathbf{METEOR}$ metrics for translation quality evaluation.
- Conducted exploratory data analysis (EDA) on sentence length distribution and vocabulary richness.
- Designed a training pipeline using **PyTorch** and **Hugging Face Datasets** for **fine-tuning** translations.

EXTRACURRICULAR

- Solved 600+ problems on various platforms like LeetCode and GFG
- Gold Medalist in Chess in Inter-Branch Chess Competition