

NITESH KUMAR

Linkedin: <https://www.linkedin.com/in/nitesh-kumar-b05051212>

Github: <https://github.com/Ni30kumr>

Email: niteshkumaar87@gmail.com

Mobile: +91-9958438491

LeetCode: <https://leetcode.com/u/Ni30kumr/>

EDUCATION

- Rajiv Gandhi Institute Of Petroleum Technology** Amethi, U.P.
Bachelor of Technology - Petroleum Engineering; CGPA: 7.00 Nov 2020-May 2024

EXPERIENCE

- Great Learning** Gurgaon
Associate Data Scientist Oct 2024 - Present
 - POC for Two Programs: MSFT-GenAI and AWS GenAI:**
 - * Addressed approximately 100 learner support queries, ensuring high satisfaction and engagement.
 - * Enhanced course content based on data insights, ensuring relevance and effectiveness.
 - Data Gathering and Data Preprocessing:**
 - * **Data Extraction:** Retrieved learners' data from Metabase using SQL queries.
 - * **Feature Engineering:** Performed feature engineering on the extracted data to enhance its usability and relevance.
 - * **Data Collaboration:** Shared the processed data with other teams for further analysis and decision-making.
- Healthiclick** Remote
Python Developer Intern Jan 2024 - 30 Aug 2024
 - Implemented RVC Deep Learning Model with Google TTS and FastAPI for Voice Cloning:**
 - * Implemented a whole pipeline for video translation of YouTube videos with emotions and proper lip sync.
 - * Integrated multiple APIs like google SST, google TTS, RVC Model **GAN based model for voice cloning**, pyaudio, assembly AI and Firebase database for hosting videos. Used Fastapi for writing backend.
 - * User will give link of video, speaker voice, language and will receive the downloadable link of video.Project Demo Link
 - FTP pipeline:**
 - * Implemented end to end pipeline with the **deployment** of call recording analysis and stored in MYSQL database for a particular date.
 - * Used FTP protocol to retrieve audio, Whisper API for transcription, GPT-3.5 model for analyzing audio, SQLALCHEMY for connection to the database, and Fastapi for writing backend.
 - * Given a date, the User can start the whole pipeline. The User can also retrieve a CSV or JSON file from the database for data between two specified dates.
 - Call GPT:**
 - * Given a customer number, my bot will call the user to remind them about their cart or to take feedback from their recent purchase.
 - * I used call recording data to prepare a dataset for fine-tuning GPT-3.5. I hosted LLaMA 3 locally using Ollama and prepared the dataset from call transcriptions. Furthermore, I used **prompt engineering** and **Langchain** to format the data set in a specific way.
 - * Integrated **fine-tuned gpt-3.5** into product.Project Demo Link

PROJECTS

- RAG-based User Bot System with FastAPI and Pinecone::**
- Built a system for users to create accounts, upload PDFs, and query personalized AI bots.
- Developed user system and **RAG**-based querying using FastAPI, SQLAlchemy, **Pinecone**, and LangChain.
- Implemented RAG-based semantic search, enabling efficient retrieval of contextually relevant information from user-specific document collections. **Tech:** Python, **Transformer**, CV2, Embedding, **Deep Learning**, Fastapi, NLP.
Github: https://github.com/Ni30kumr/RAG_pipeline_with_langchain.git

SKILLS SUMMARY

- Languages:** Python, C++, **SQL**
- Library:** Scikit, NLTK, TensorFlow, Matplotlib, Pandas, Numpy, Opencv, **Pytorch**, **Hugging Face**, **Langchain**
- Tools:** **FastAPI**, **Docker**, **MYSQL**, **AWS**, **Git**, **NLP**, MongoDB
- Platforms:** Windows, VS code, Jupyter Notebook, Google Colab, **Linux**
- LLMs:** GPT Family, **Falcon**, **Mistral**, **Llama**, **MPT**

HONORS AND AWARDS

- Top 2 percentile in JEE ADVANCE qualified JEE Mains - Oct-2020
- Merit-cum mean scholarship at RGIPT 2020-2024
- Qualified Data Science and Artificial Intelligence exam of GATE 2024
- Got selected for EAGE Mentor-mentee program Dec 2023