Arun Singh Negi

Machine Learning

✉ arunpycodecmaster@gmail.com 📞 8077840135 📍 Delhi, India in LinkedIn ○ Github 🔗 Portfolio

PROFILE

Data science professional with 3 years of experience in technical support, specializing in real-time problem-solving in machine learning and deep learning. Proven track record in developing scalable code in a cloud environment. Keen to stay current with industry trends for efficient, impactful data-driven solutions.

SKILLS

statistics — *Descriptive statistics | Inferential statistics*, **Language** — *Python,SQL,HTML,CSS,JS*, **Machine Learning** — *Supervised | Unsupervised Learning*, **Natural Language Processing-NLP** — *RNN | LSTM | GRU | Word2vec | Transformers | Encoder-Decoder*, **Generative AI** — *LLM | Vector Databases-Pinecone,FAISS,Chroma | llama-2,OpenAI,Mistral,Groq,Gemma*, **MLOPS** — *GitHub | Docker | AWS ECR | AWS Bucket | AWS EC2 | CI-CD*, **Data Analytics** — *Pandas, Seaborn, Matplotlib*

PROFESSIONAL EXPERIENCE

INEURON INTELLIGENCE PRIVATE LIMITED
Internship

Feb 2023 – Jan 2024
Bengaluru, India

- I worked on end-to-end projects understanding the problem statement and designing an architecture for optimal solution.
- Conducted exploratory data analysis to comprehend problem statements, employing analytical tools including Pandas, Matplotlib, Seaborn, and Power BI.
- Developed scalable code and integrated seamless pipelines with databases for various projects, including thyroid disease detection, object detection, text summarization, and more.
- Applied diverse algorithms and incorporated pre-trained models in projects to enhance performance metrics.
- Utilized Docker and GitHub to deploy projects on cloud platforms, establishing a continuous integration and continuous deployment (CICD) pipeline.
- Documentation of High-level design and Low-level design with proper workflow.

Teleperformance
Team Lead

May 2022 – Jul 2023
Gurugram, India

- As a Team Lead at Teleperformance, I have successfully led a team of 18 specialists and ensured the delivery of exceptional customer service and support.
- Real-time resolution of client issues regarding data manipulation and monitoring live feed to dialer.
- Worked on high-level escalation of the English world and APAC region on software and hardware-related issues.
- Worked on issues related to software installation through Jenkins, AD manager, TP console, Azure AD, etc.
- Ensuring compliance with company policies and procedures, and maintaining a positive and professional work environment.

Teleperformance
Technical Support

Jul 2020 – May 2022
Gurugram, India

- As a Technical Support Specialist at Teleperformance, I provided technical support to customers and resolved their IT-related issues.
- Provided real-time assistance to clients for Genesys dialer using SQL queries.
- Provided support to client regarding active directory issues like user creation, authentication issues.
- use cmd to resolve remote software issues and perform diagnosis.
- Documented and maintained records of customer interactions and issue resolutions in the service desk ticketing tool.

PROJECTS

Chatbot 🔗
Gen AI

  Github:https://github.com/Arun02DS/chatbot_medical.git
  -**Tech Stack** - Python,Pinecone vector db,Langchain, Meta lamma2, Flask
  -Utilized pre-trained LLAMA2 model to deploy Flask-based application.
  -Trained model on medical encyclopedia data, enabling it to answer related queries proficiently.
  -Employed text embedding with overlap processing, stored in Pinecone for efficient retrieval.

Text Summarization 🔗
Summary of Text (NLP project)

  GitHub: https://github.com/Arun02DS/Text-summarization.git
  - **Tech Stack**- Python, Huggingface Transformers, Pegasus-cnn_dailymail,Github, FastAPI, Docker, AWS
  - I implemented a text summarization solution by incorporating a pre-trained model within an end-to-end pipeline.
  - The pipeline encompasses data collection, transformation, training, and evaluation stages.
  - An application has been created to facilitate real-time implementation checks, whether offline or on the cloud.
  - After the process of dockerization, the image has been successfully deployed to the AWS cloud.

Thyroid Detection Problem ✎
Classification Problem
  GitHub: https://github.com/Arun02DS/Thyroid_detection.git
  – **Tech Stack** - Python, MongoDB, GitHub Actions, Docker, S3 bucket, AWS EC2, AWS ECR, Apache Airflow.
  – I designed an ML pipeline with six key components such as data ingestion,data validation, data transformation, etc. emphasizing robustness through integrated logging and thorough exception handling in the codebase.
  – Initiated data ingestion by extracting data from MongoDB and incorporating it into the pipeline
  – Created an image using Docker and deployed it to the AWS cloud through GitHub Actions.
  – AWS components come into play as the artifacts are stored in an S3 bucket, images are housed in the ECR repository, and the execution of code takes place within an EC2 instance integrated with Apache Airflow.

MCQ Generator ✎
Gen AI
  GitHub: https://github.com/Arun02DS/Langchain_generator.git ✎
  – **Tech Stack**- Python, Langchain, OpenAI, Streamlit
  – An streamlit application was trained on Text data and responce output as MCQ questions based on that text was developed.
  – It uses OpenAI "gpt-3.5-turbo" model to build a sequential chain with prompt and input parameters.
  – Model get trained on input text using given prompt give responce as a JSON text.

EDUCATION

Masters of Technology   Oct 2013 – Jul 2015
Govind ballabh pant university of agriculture and technology   Pantnagar, Uttrakhand, India

Bachelor of Technology   Aug 2007 – Apr 2011
Uttrakhand technical university   Dehradun,Uttrakhand, India

CERTIFICATES

Full Stack Data Science Bootcamp ✎ *— May 2022 - Oct 2023* |
Oracle Cloud Infrastructure 2023 Certified Data Science Professional ✎

DECLARATION

I hereby declare that the details furnished above are true and correct.

Arun Singh Negi