PRAVEEN KUMAR

Jr. Data Scientist

2.7 Years Experience

Mobile: +919006336740

Email: praveenku32k.be@gmail.com

GitHub: Praveenku32k (Praveen Kumar) (github.com)

LinkedIn: LinkedIn

Experience Summary

- Data Scientist with 2.7 Years of experience in analyzing a large dataset, and developing Machine Learning Models for various problem statements using various applications of Computer Vision, Large Language Models, Generative AI, Prompt Engineering and LangChain.
- I have Fine-Tune Mistral, AllMiniLM-L6-V2,OpenAI, Llama-2, Zephyr ,Qwen1.5 using Lora and Qlora technique. Exploring DPO and RLHF.
- I hold a Post Graduate Program in Computational Data Science from Case Western Reserve University, USA and a Bachelor of Engineering in Computer Science and Engineering from SJB Institute of Technology.
- Experience in developing Python scripts using various libraries such as TensorFlow, scikit-learn, Numpy,Pandas, Matplotlib, Keras API, and OpenCV for data analysis and numerical computations.
- Involved in creating various regression and classification algorithms by using various scikit-learn libraries such as Linear Regression, Decision Trees, and Random Forest.
- Capable of developing deep-learning models using multiple frameworks such as TensorFlow, Keras, PyTorch.

Technical Skills

- **Language:** Python, C++
- **GenerativeAI**: LLM-Vicuna-7b, Llama-7b, Mistral-8b, Falcon-7b, ChatGPT, OpenAI, LORA/QLORA, Gimini, M2M, Fine-Turning, RAG, DPO, Evaluation(Rouge, Ragas, BLUE), Embedding(AllMiniLM-L6-V2,OpenAIEmbedding).
- **LLM Frameworks**: LangChain, LlamaIndex, Keras, TensorFlow, PyTorch
- **Vector Database:** Redis, Pinecone, Chromadb, Faiss, Neo4j
- **Python Libraries**: Pandas ,Numpy ,Sklearn, matplotlib, seaborn, Scikit Learn
- **Machine Learning**: Statistics, Statistics Modeling, Linear Regression, Logistic Regression, KNN, SVM, Decision Tree, Random Forest(Bagging, Boosting, Stacking), PCA, Clustering.
- **Deep Learning Advance**: Perceptron, Activation Function, Loss function, Optimizer, Regulation
- **Computer Vision**: LeNet-5, AlexNet, VGG, GoogleNet,ResNet, Object detection, Object Classification, YOLO, Segmentation, Generative Adversarial Neural Netwok (GAN), SSD.
- **Natural Language Processing**: NLTK, Spacy, Standford,LSTM, GRU, Encoder and Decoder, Transformer, Bert, BART, ChatGPt, Attention Is All You Need(Transformer))
- **DevOps**: Git, GitHub, Docker, Jenkins,Linux
- **Cloud**: **AWS** Sagemaker,EC2,S3,BedRock,Lambda,**GCP**-Basic, **Azure**-basic.
- **Others**: Data Structure, Flask, Gradio, Excel,Postgres, Redis.

Organizations

- **RevealIT Software Pvt. Ltd - Jr. Data Scientist** (06/2023-Present)
- **TuringMinds.Ai/Insofe/Soothsayer Analytics** -Data Scientist (02/2022 -4/2023)

  Note:- Trainee followed by Full-Time in SoothSayer Analytics.
- **WIPRO** -Full Stack Developer (Internship) **(**03/2022 - 05/2022)

Academic

Level	Year of Passing	University / Board	Percentage/CGPA/Completion
Postgraduate Program in Computational Data Science	2023	Case Western Reserve University, USA	Certification Completed
Bachelor of Engineering	2022	SJB Institute of Technology, Bangalore	8.1

Projects

- **Data Normalizer** LLM **(Present -on Going Project)**
  - ✓ This project is about Normalizing the data coming from different source like UCMDB, ServiceNow, Flexera etc. It have 10 Millions of data to do Preprocessing.
  - ✓ Data Normalizer do similarity search on Redis, Uisng Redis Vector Similarity search.
  - ✓ I have used embedding model 'all-MiniLM-L6-v2' model. I have used Redis as vector store.
  - ✓ This project used Redis, Postgres db,Docker, HuggingFace model and Flask Api.
  - ✓ I have worked on Fine-Tuning of 'all-MiniLM-L6-v2' model.
- **Vicuna-13b(Chatbot)+ Private Pdf's Chatbot (LLM)**
  - ✓ I have used vicuna-13b model from HuggingFace.I have used `all-MiniLM-L6-v2` for Embedding .
- **Attendance System for Face Recognition.**
  - ✓ The software was to manage the attendance system. I have used images as a data set for training the model and integrated camera. It also has a reportingsystem to keep track of work assigned and the performance of employees.The accuracy achieved was 98.89%
  - ✓ Various algorithms used such as KNeighborsClassifier, RandomForestClassifier,PCA, ANN-ReLU(Activation Function)
- **Package: App-ANN-Project-Praveenku32k (Open Source)**
  - ✓ Deep understanding of Perceptron using forward and backward propagation and trying to use different Activation Functions.
  - ✓ This package is to check the different functions of Artificial Neural network.
  - ✓ I have developed PyPI (The Python Package Index).pip install App-ANN-Project-Praveenku32
  - GitHub Link: Praveenku32k/App_ANN_Project: App_ANN_Project (github.com)

Certificates

- Machine Learning and Deep Learning Masters issued by iNeuron(08/2021-07/2022)
- Deep Learning , Computer Vision and NLP Master issued by iNeuron(08/2021-08/2022)
- Langchain, Pinecone, OpenAi: Build Next-Gen LLM App by Udmey(06/2023-07/2023)
- Mastering Generative AI with OpenAI, Langchain,and LlamaIndex (02/2024-06-2024)