PRAVEEN KUMAR   **Email:** praveenku32k.be@gmail.com

Jr. Data Scientist   **GitHub:** Praveenku32k (Praveen Kumar) (github.com)

Mobile+91-9006336740   **LinkedIn:** LinkedIn

Professional Summary

- Data Scientist with 2.9 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 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.

- 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.

Skills

- **Language - (***Python* **|** *C++***)**

- **Generative AI-**(*Langchain* **|** *OpenAI API* **|** *Prompt Engineering* **|** *RAG***|***GPT3* **|** *Llama-7B* **|***T5***|** *BERT* **|** *MultiModel* **|** *Agentic-AI* **|** *Embedding Models*)

- **Vector Database**-(Redis **|** Pinecone **|** Chromadb **|** Faiss **|** Neo4j)

- **Python Libraries-(***Pandas* **|** *Numpy* **|** *Sklearn* **|** *matplotlib* **|** seaborn **|** Scikit *Learn* **|** *Keras* **|** *TensorFlow* **|** *PyTorch* **|** *Hugging Face Transformers*)

- **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 -(***Convolutional NeuralNetwork* **|** *Object Classification* **|** *Object Detection* **|** *Object Segmentation*)

- **Natural Language Processing-(***NLTK* **|** *Spacy* **|** *Standford* **|***RNN* **|***LSTM* **|** *GRU* **|** *Encoder and Decoder* **|** *Transformer* **|** *Bert* **|** *BART* **|** *Chat GPT*)

- **DevOps-(***Git, GitHub* **|** *Docker* **|** *Jenkins* **|** *Linux*)

- **Cloud**: **AWS-(***Sagemaker,EC2,S3,BedRock,Lambda* **|** **GCP**-*Basic* **|** **Azure**-*basic*)

- **Others-(** *Data Structure* **|** *Flask* **|** *Gradio* **|** *Excel* **|** *Postgres* **|** *SQL*)

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)

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.

- **Private Pdf's Chatbot with Llama-2-7b**
- Data Preparation: Leveraged understanding of existing procedures and workflows in coordination with the client to prepare data for ingestion.
- Collaborated with the client to prepare ground truth data, crucial for the evaluation of Language Model (LLM) performance.
- Constructed an initial pipeline with candidate LLMs (chatGPT3.5-turbo, Llama-2-7b, FLAN-T5-XL,Google-PALM2) and meticulously documented their performance.
- Applied various prompt-engineering techniques (Zero-shot, Few-shot,Chain-Of-Thought prompting, etc.) and logged the results for evaluation using MLFlow
- Implemented Retrieval Augmented Generation (RAG) techniques to fetch relevant information .
- Conducted research on suitable evaluation metrics (ROUGE-L,BLUE Score , etc) and implemented them to evaluate the results.

- **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)-Personal**
- 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

Education

- Postgraduate Program in **Computational Data Science** from Case Western Reserve University, USA (Certification) **2022-2023**
- Bachelor of Engineering in **CSE** from SJB Institute of Technology, Bangalore, **2018-2022**

Certification

- Machine Learning and Deep Learning Masters issued by iNeuron(08/2021-07/2022)
- 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)