

**PRAVEEN KUMAR**

**Jr. Data Scientist**

**2+Years Experience**

**Mobile: +919006336740**

**Email: [praveenku32k.be@gmail.com](mailto:praveenku32k.be@gmail.com)**

**GitHub: [Praveenku32k \(Praveen Kumar\) \(github.com\)](https://github.com/PraveenKumar32k)**

**LinkedIn: [LinkedIn](#)**

**Linktree: [@praveenku32k.be](https://praveenku32k.be) | [Linktree](#)**



### Experience Summary

- Data Scientist with 2+Years of experience in analyzing a large dataset, and developing Machine Learning Models for various problem statements using various applications of Computer Vision, NLP(LLM MODELS),Generative AI, Prompt Engineering, Langchain and Llama-Index.
- Experience in developing Python scripts using various libraries such as TensorFlow, Scipy, Numpy,Pandas, Matplotlib, Keras API, and OpenCV for data analysis and numerical computations.
- Involved in Data Preprocessing Techniques for making the data useful for creating Machine Learning models.
- Involved in creating various regression and classification algorithms by using various sklearn libraries such as Linear Regression, Decision Trees, and Random Forest.
- Capable of developing deep-learning models using multiple frameworks such as TensorFlow, Keras, PyTorch, and Detectron2.

### Technical Skills

#### Technical Skills Set:

- Language(Python(OOP's Concept)), C++ with Data Structure.
- LLM(Vicuna-7b, 13b,Llama,Fastchat,Mistral-7B-x , ChatGPT, BARD, Llama-7b, ,M2M, FineTurning of Llama2,Gemma, Gemini-pro, Embedding Models,)
- Langchain(Chain ,Agent, Prompt,Template, HuggingFace) ,Llama-Index, RAG, GAG - Evaluation like Ragas, Blue. Database like Pinecone, chromadb, fassi, weavite, Redis.
- Frameworks(Keras, TensorFlow, PyTorch), Experience in GPU
- Python Libraries(Pandas, sklearn, numpy, matplotlib, seaborn, Scikit Learn, and Detectron2 etc.)
- Machine Learning(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 ( Basic,LeNet-5, AlexNet, VGG-16,19,GoogleNet,ResNet,Object detection, Object Classification, YOLO-V1,2,3,4,5,X,R,Segmentation, GAN,SSD)
- Natural language processing(NLTK, Spacy, Stanford), LLM, LSTM, GRU, Encoder and Decoder, Transformer, Bert, BART, ChatGpt-V1, V2, Chat GPT(Basic), Attention Is All You Need(Transformer))
- DevOps(Git, GitHub ,Docker, Jenkins)
- Cloud(AWS(Sagemaker,EC2,),Heruko,VertexAI,Weaviate,)

## Projects

- **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.
  - ✓ First I have used langflow for the flowchart. Based on flowchart made this project.
  - ✓ This project is about chat with multiple private pdf with Chromadb vectorstore and pinecone.
  - ✓ I have used langchain to create this project.
- **Fine-Tuning of Llama2 (OpenSource Available to HF)**
  - ✓ I have fine-tune Llama2 model which is publicly available in Huggingface.
  - ✓ Link for this Model: <https://huggingface.co/praveenku32k/Llama-2-7b-chat-new-qA-finetune>
- **Telegram Chat Bot with Gemini-pro model**
  - ✓ TO build this model I have aiogram library which is publicly available to built anychatbot like whatsapp, Telegram.
  - ✓ I have use chatgpt-3.5-turbo first and then later I change to Gemini-pro.
  - ✓ From This project you can chat with gemini model from your Telegram only.
- **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 \(github.com\)](https://github.com/praveenku32k/App-ANN-Project)
- **YOLO-V4 for Object Detection**
  - ✓ This project is used to detect object using computer vision (YOLO concept).
  - ✓ The data set which is given is 15k for training purposes and 5k for testing purpose.
  - ✓ I have used ResNet-34 and DarkNet-19 as a Architecture.
  - ✓ I got 94% of accuracy which is better than the other architecture.
  - ✓ Finally, I was able to detect small and big object both.
- **Data Normalizer (Present -on Going Project)**
  - ✓ This project is about Normalizing the data coming from different source like UCMDB, ServiceNow, Flexera etc.
  - ✓ Data Normalizer do similarity search on Redis, Using Redis Vector Similarity search.
  - ✓ I have used embedding model '**all-MiniLM-L6-v1**' model.
  - ✓ I have used Redis as vector store and Postgre as DataBase. Later we are planning to move to pinecone.
  - ✓ This project used Redis, Redis Cache, Postgres db, HuggingFace model and Flask Api.

## Organizations

- **WIPRO (Internship)**
  - ✓ **Full Stack Developer** (03/2022 - 05/2022)  
**Project:** Online Exam System  
**Technology used:** DotNet Framework, HTML  
**Database:** SQL  
**Summary:** This application was to develop an Online exam system using .net Framework. This takes the exam online and only admins are allowed to modify the questions. The main thing about this application is admin can monitor the exam. We have used Integrated cameras to monitor the exam.
- **TuringMinds.Ai/ Soothsayer Analytics**
  - ✓ **Data Scientist (04/2022 -4/2023)**
  - ✓ I started as a Trainee Data Scientist and Promoted as Data Scientist. The Project which I have worked on project which are mention above.
- **RevealIT Software Pvt. Ltd.**
  - ✓ **Jr. Data Scientist (06/2022-Present)**
  - ✓ I am working on Data Normalizer Project.
  - ✓ Working with Large Language Models and Langchain to create ChatBot with Vicuna.

## Certificates

- Python for Beginner” conducted by Electronics & ICT Academy, IIT Roorkee with cloudxlab. (01/2020 - 03/2020)
- “Python for Data Science” issued by IBM (05/2020 - 06/2020)
- "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)

## Academic Profile

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
12th	2017	CBSE	69.6%

## Personal Details

Fathers Name: Sunil Kumar Yadav

Gender: Male

Date of Birth: 17-03-2000

Languages: English and Hindi

Address : Vill-Pipradih, P.O-Saranda, P.S-Birni, Dist-Giridih, Jharkhand

Present Addr. - Bangalore(whitefiled)